Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers

Gilles LC Baud, Aiswarya Prasad, Kirsten M Ellegaard, Philipp Engel

Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland

Correspondence:

Prof. Philipp Engel

Department of Fundamental Microbiology

University of Lausanne, CH-1015 Lausanne, Switzerland

Tel.: +41 (0)21 692 56 12

e-mail: philipp.engel@unil.ch

Keywords: metagenomics, gut microbiota, strain diversity, honey bee, symbiosis, social insects

Abstract

Strain-level diversity is widespread among bacterial species and can expand the functional potential of natural microbial communities. However, to what extent communities undergo consistent shifts in strain composition in response to environmental/host changes is less well understood. Here, we used shotgun metagenomics to compare the gut microbiota of two behavioral states of the Western honeybee (*Apis mellifera*), namely nurse and forager bees. While their gut microbiota is composed of the same bacterial species, we detected consistent changes in strain-level composition between nurses and foragers. Single nucleotide variant profiles of predominant bacterial species clustered by behavioral state. Moreover, we identified strain-specific gene content related to nutrient utilization, vitamin biosynthesis, and cell-cell interactions specifically associated with the two behavioral states. Our findings show that strain-level diversity in host-associated communities can undergo consistent changes in response to host behavioral changes modulating the functional potential of the community.

Introduction

Microbial communities are highly dynamic. They undergo compositional changes in response to environmental factors, which can profoundly influence their functions. For example, antibiotic treatment can disturb the human gut microbiota, altering its effect on the host metabolism and immune system, and facilitating the invasion of pathogens [1]. Such microbiota shifts are difficult to predict as they often occur irregularly and in different contexts. However, when linked to recurrent or predetermined environmental changes, microbial communities can undergo consistent compositional transitions, providing excellent opportunities to study ecological interactions governing community shifts. For example, bacterioplankton communities in the ocean that depend on sunlight intensity [2-4], or gut microbial communities of animals that have a seasonal diet, or that hibernate during the winter months [5-8], turnover as a function of season. Similarly, the assembly of the human gut microbiota after birth follows a predictable pattern characterized by the succession of different microbial taxa during the first few months of life [9, 10].

Compositional changes associated with the turnover of microbial communities have mostly been studied at the genus- or species-level using 16S rRNA gene analysis. However, natural microbial communities usually harbor a large extent of strain-level diversity, with strains of the same species carrying distinct functions. In the human gut, strain-level differences can influence the metabolism of dietary compounds [11, 12] or drugs [13], or modulate their interactions with the host immune system [14]. Moreover, several recent studies have suggested that strains of the same species can exhibit different dynamics and govern the eco-evolutionary dynamics in microbial communities [15-18]. However, if natural microbial

56 communities undergo consistent compositional transitions at the strain-level in response to
57 environmental changes is less well understood.

58 Honeybees (*Apis mellifera*) can be useful models for studying strain-level dynamics in host-
59 associated microbial communities. They are eusocial animals that live in large colonies
60 composed of a single queen, a few hundred male bees, and thousands of sterile female
61 worker bees. They harbor a relatively simple, but highly specialized microbial community in
62 their gut, composed of <10 bacterial phylotypes (i.e. bacteria sharing >97% sequence identity
63 in the 16S rRNA gene) [19, 20]. Most of these phylotypes have diverged into sequence discrete
64 populations (sequence clusters of >95% ANI; hereafter referred to as species), each harboring
65 a high degree of strain-level diversity and variation in functional gene content [21-26].

66 Using shotgun metagenomics, we have previously shown that while phylotypes, and species
67 within phylotypes, co-exist in individual bees, strains within species segregate, suggesting that
68 each worker bee harbors a distinct community at the strain-level [22]. These differences in
69 strain-level composition could not be attributed to any variation in age, time of sampling, or
70 colony affiliation. However, in our previous study, only a few colonies were sampled, and bees
71 were exclusively sampled from within the hive between September and February (i.e. autumn
72 and winter) raising the possibility that most microbiota samples originated from nurse bees
73 or long-lived winter bees. During the summer months, worker bees undergo an age-related
74 behavioral maturation [27-29]. Young nurse bees feed on pollen and fulfill tasks inside the
75 hive related to brood care, nest building, and food processing. However, when 2-3 weeks old
76 nurses transition to forager bees, which feed on honey, and collect pollen and nectar outside
77 the hive. Besides the dramatic change in dietary intake, this behavioral maturation is
78 accompanied by a series of physiological changes, such as endocrine secretion, fat body

79 metabolism, and brain activity [30, 31]. Community profiling based on 16S rRNA gene
80 sequencing has shown that the nurse-to-forager transition is associated with a consistent
81 change in gut microbiota composition [32, 33]. While the microbiota of both bee types is
82 dominated by the same bacterial phylotypes, the composition of the community shifts with a
83 marked decrease in the absolute abundance of most phylotypes in forager versus nurse bees.

84 Here, we tested if the gut microbiota of honeybees during the transition from nurse to forager
85 bees also undergoes consistent compositional changes at the strain-level. To this end, we
86 analyzed shotgun metagenomics data from nurse and forager bees sampled across 15
87 colonies in Western Switzerland. Our results show that the two behavioral states harbor
88 distinctive bacterial communities across colonies, characterized by consistent shifts in strain
89 composition and functional gene content of major bacterial species. These findings show that
90 strains are ecologically distinct units that undergo a predictable ecological transition when
91 bees turn from nurses to foragers. We speculate that this microbiota shift is driven by nutrient
92 availability disparities or physiological changes in the gut of foragers relative to nurses.

93

Results

Bacterial loads and species-level community composition differ between nurse and forager bees

To investigate compositional changes with genomic resolution in the honeybee gut microbiota in response to behavioral state, we collected metagenomic samples from larvae-feeding nurses and pollen foragers (**Supplementary Table 1).** Samples consisted of pooled honeybee guts to ensure representative sampling of the strain-level diversity for each colony. In total, we sampled 15 colonies from five different sampling locations (three each), which allowed us to account for possible variation between sites and colonies, not related to behavioral state.

Due to the different dietary habits of the two behavioral states, the hindguts of nurses were usually pollen-filled, as indicated by the enlarged gut size, and the opacity and yellow color of the gut content. In contrast, the hindguts of foragers were smaller and contained mostly nectar based on their translucent appearance. Accordingly, nurse guts were on average 2.9x heavier than their forager counterparts (p=3.3e-09, Wilcoxon sign rank test, **Supplementary Figure 1)**.

To estimate the absolute abundance of bacteria relative to the amount of total DNA in each pooled gut sample, we performed qPCR on genomic DNA targeting the bacterial 16S rRNA gene. This revealed that the bacterial loads in the forager samples were on average 2.6x lower than the ones in the nurse samples (p=1.83e-04, Wilcoxon signed-rank test, **Figure 1A**), confirming previous results by Kešnerová et al. (2020).

Shotgun metagenomics sequencing resulted in 58.4 – 76.7 million reads per sample. To compare the bacterial community composition of nurses and foragers, the metagenomic

117 reads were mapped to a reference database of isolate genomes of strains from the gut
118 microbiota of social bees, as in Ellegaard et al. (2020) and Ellegaard and Engel (2019). The
119 non-redundant database was composed of 198 isolate genomes and encompassed all major
120 community members described to be present in the honeybee gut microbiota. Between 19.8
121 and 57.4 % of all metagenomic reads mapped to the bacterial reference database. Of the
122 remaining reads, 30.6 - 76.0 % mapped to the host genome of *A. mellifera,* and only 4.2 - 20.8
123 % of the reads could not be assigned (**Supplementary Figure 2A**). Forager samples overall had
124 a lower proportion of reads mapping to the bacterial database than the nurse samples
125 (forager average 35.6%, nurse average 50.8%, p=6.1e-05, Wilcoxon signed-rank test), which
126 is in line with the difference in bacterial loads between the two behavioral states. Taxonomic
127 profiling with mOTU2 [34] revealed that a large fraction of the unmapped reads originated
128 from bacteria, especially from *Bartonella*, *Snodgrassella*, and *Lactobacillus* suggesting that
129 some of the diversity of these core microbiota members is not covered in our reference
130 database. However, also some accessory community members were detected among the
131 unmapped reads such as *Hafnia*, *Klebsiella*, or *Enterobacter*, especially in nurse samples
132 (**Supplementary Figure 2B**).

133 We next determined the relative and absolute abundance of the major phylotypes and
134 species in each sample based on the read coverage of single-copy core genes and the total
135 number of 16 rRNA gene copies determined by qPCR. We will refer to different species within
136 phylotypes using the previously established SDP (sequence-discrete population) numbering
137 [22, 35], for reasons of readability but also because only for a subset of them proper species
138 names have been established to date. A list with all SDPs and corresponding species names
139 can be found in **Supplementary Table 2**. Both the phylotype and species-level composition
140 were comparable to the community profiles detected in previous shotgun metagenomic

141 analyses based on individual bees [22, 35]. Firmicutes (Bombilactobacilli and Lactobacilli) and
142 Bifidobacteria dominated the community, and for the phylotypes Bombilactobacillus,
143 Lactobacillus, Bifidobacterium, and Gilliamella, different species were detected in all samples
144 (**Figure 1B, Supplementary Figure 3**). Consistent with previous findings that social bees
145 harbor host-specific communities [35, 36], most reads mapped to the genomes of bacterial
146 isolates from *A. mellifera*. Only a few reads (0.15% - 3.1%) in nine samples mapped to bacteria
147 from other bee species (including Lactobacilli isolated from *Apis cerana,* and Gilliamella
148 isolated from bumble bees and *Apis cerana*, **see Supplementary Table 3**), which may be due
149 to spurious mapping.

150 Non-metric multidimensional scaling (NMDS) using Bray-Curtis dissimilarity of the bacterial
151 relative abundances separated nurse from forager samples at both the phylotype-
152 (PERMANOVA, Host – R^2=0.40279, p-value=0.001) and the species-level (PERMANOVA, Host
153 - R^2=0.45201, p-value=0.001, **Figure 1C and 1D**). Samples also clustered by location
154 (**Supplementary Figure 4**). However, it is important to note that our study was not designed
155 to test for this, as only three colonies were sampled per location with two samples per colony.

156 In terms of absolute abundances (as determined by multiplying the proportions of each
157 community member by the total bacterial biomass in each sample, see methods), most
158 community members were less abundant in foragers than in nurse bees (**Supplementary**
159 **Figure 5**), a direct consequence of the lower total bacterial biomass present in the gut of
160 foragers (**Figure 1A**). However, there were clear differences in the strength of these changes
161 between community members, even among species of the same phylotype (**Figure 1E**). For
162 example, the total abundances of Lactobacilli Firm5_1 and Firm5_4 changed to a lesser extent
163 (average log2 fold change = -1.45 and -0.71) than those of Firm5_2 and Firm5_3 (average log2

164 fold change = -2.82 and -2.99) in foragers relative to nurses. This explains the consistent shift
165 towards a less even species composition of this phylotype in foragers compared to nurses as
166 determined with the Shannon diversity index (Wilcoxon signed-rank test, p-value=8.8e-06,
167 **Supplementary Figure 3**). Likewise, Bifido_1.2 was more affected (average log2 fold change
168 = -2.32) than any of the other species of the Bifidobacterium phylotype (average log2 fold
169 change across species of the Bifidobacterium phylotype = -0.37) by the behavioral state
170 transition of the host (**Supplementary Figure 3**). Notably, four community members (Bapis,
171 Snod_1, Com_1 and Lkun) did not change in absolute bacterial load between nurses and
172 forager samples, suggesting that they are not affected in their abundance by the behavioral
173 transition.

174 In summary, although the gut microbiota of forager and nurse bees were composed of the
175 same community members, we found consistent differences in the absolute abundance of
176 many of them, not only at the phylotype-level as previously shown [32, 33], but also at the
177 species-level. Moreover, some species of a given phylotype were more strongly affected than
178 others suggesting species-specific responses to the nurse-to-forager transition.

179

180 **Consistent differences in strain-level composition between nurse and forager bees**

181 Previous studies have shown that most honeybee gut microbiota species harbor high levels
182 of strain diversity and that the strain composition differs between individual bees [22, 35]. To
183 assess strain-level differences between nurses or foragers, we mapped the metagenomic
184 reads against a reduced reference database containing one genome per species and
185 quantified the number of polymorphic sites in the reads mapped onto the single-copy core
186 genes of these reference genomes.

187 We detected between 1.85 and 11.19% polymorphic sites per species per sample, with
188 Firm4_2 showing the lowest, and Firm5_4 the highest average proportion of polymorphic
189 sites per sample (**Figure 2A**). These differences in proportions of polymorphic sites among
190 community members are comparable to those previously found by Ellegaard et al. (2020) and
191 Ellegaard and Engel (2019). To assess the total amount of strain-level diversity detected per
192 species across all samples, we calculated cumulative proportions of polymorphic sites as a
193 function of the number of bee samples. Firm4_2 showed the lowest (4.91%) and Bifido1_1
194 the highest (12.8%) level of total diversity across all samples (**Figure 2B**). In all cases, we
195 observed that the slope of the curves rapidly decreased with the addition of new samples,
196 indicating that many of the polymorphic sites are shared across samples.

197 When comparing the two behavioral states, we found that five of the 17 analyzed species
198 showed higher strain-level diversity in nurses than in foragers (Wilcoxon signed-rank test,
199 Firm4_1: q-value=4.20e-3, Firm5_1: q-value=6.45e-3, Firm5_2: q-value=1.08e-2, Firm5_3: q-
200 value=3.56e-4, Gilli_1: q-value=6.45e-3) whereas only two species showed lower strain-level
201 diversity in nurses than in foragers (Wilcoxon signed-rank test, Gilli_2: q-value=2.91e-2,
202 Snod_1: q-value=2.91e-2, **Figure 2A**). The forager samples had on average 2.6 times lower
203 bacterial load than nurse samples, but on average only 1.5 times smaller number of reads per
204 sample. This means that a larger fraction of the total bacterial population was analyzed for
205 forager relative to nurse samples, a difference that could potentially affect diversity
206 estimates. However, no significant correlation was found between the proportion of
207 polymorphic sites detected and the read coverage of our samples (**Supplementary Figure 6**),
208 suggesting that we had sufficiently deep coverage across the study.

209 Next, we assessed the difference in strain profiles between forager and nurse samples by
210 comparing the distribution of single nucleotide variants (i.e. alleles) at the detected
211 polymorphic sites for each species across samples. To do so, we computed the Jaccard
212 distances between all sample pairs based on the number of shared polymorphic sites.
213 Principal coordinate analysis of these similarity indices resulted in clustering of the samples
214 by host behavioral state for 11 of the 17 analyzed species (PERMANOVA, p-value < 0.05,
215 **Figure 2C**), including all species of Bifido_1 and Firm5, except for Bifido_1.5 and Firm5_4,
216 Firm4_1, Gilli_1, Gilli_2, and Snod_1. Notably, we also observed clustering by location in the
217 principal coordinate analysis (PERMANOVA, p-value < 0.05, **Supplementary Figure 7**).
218 However, as already stated before, our study was not designed to test for the effect of
219 location as two samples per location came from the same colony.

220 In summary, our analysis shows that nurse bees and foragers have distinct strain profiles for
221 many bacterial species both in terms of the amount of diversity and type of variants
222 detected.

223

224 **Intraspecific differences in functional gene content between nurse and forager bees**

225 Previous studies have shown that most species of the honeybee gut microbiota harbor large
226 accessory gene pools and that individual bees vary in terms of the functional gene content
227 encoded in a given species [22]. Because nurses and foragers have different physiology and
228 diet, we thus hypothesized that the observed changes in strain-level composition between
229 the two behavioral states may be driven by functional differences among strains.

230 To identify strain-specific genes that may be enriched in one behavioral state or the other,
231 we compared the read coverage of all genes assigned to a given species. Strain-specific genes
232 should have a lower read coverage than the single-copy core genes of a given species; and
233 those that predominate in one behavioral state relative to the other should show a consistent
234 coverage discrepancy between forager and nurse samples. To identify such genes, we
235 assembled all reads not mapping to the host genome and binned the resulting contigs into
236 bacterial species bins using our reference database (see methods for details and
237 **Supplementary Figure 8**). Then, we clustered the ORFs encoded in the binned contigs into
238 species-specific ortholog groups (OGs) and identified those OGs that showed consistent
239 coverage discrepancy between forager and nurse samples. In total, 81.7% (65.3 – 90.9% per
240 sample) of all assembled bases could be binned and 83.4% (73.2 – 86.1% per sample) of the
241 non-host reads mapped to the binned contigs indicating that we had included most of the
242 bacterial diversity present in the samples. Out of 45,863 species-specific OGs, 1,551 revealed
243 significant differences between the nurse and forager samples (q-value < 0.05, Wilcoxon
244 signed-rank test, and absolute log2 fold change > 1, **Supplementary Dataset 1**), which
245 indicated that these genes were more abundant in one host behavioral state versus the other.
246 Notably, the number of differential abundant OGs substantially varied between species
247 ranging from 15 to 231 (**Figure 3A, Supplementary Dataset 1**). For example, in Bifido_1.1, we
248 detected very few consistent gene content differences between foragers and nurses (6 and 9
249 OGs, respectively). This species had the highest proportion of polymorphic sites across
250 samples, but showed no clustering of the samples by behavioral state based on the number
251 of shared polymorphic sites (**Figure 2C**). The largest differences in gene content between
252 nurses and foragers were found for three Lactobacillus Firm5 species (Firm5_2: 184 OGs,
253 Firm5_3: 131 OGs, Firm5_4: 177 OGs), two Gilliamella species (Gilli_1: 64 OGs, Gilli_2: 161

254 OGs), Snod_1 (197 OGs), and Bifido_1.2 (159 OGs). Consistent with these findings, all of these
255 species, except for Firm5_4, also showed clustering by behavioral state based on the number
256 of shared polymorphic sites (**Figure 2C**). Interestingly, a larger fraction of the state-specific
257 OGs were found in foragers than in nurses (in total 1019 versus 445 significantly different
258 OGs, respectively).

259 To determine the functions encoded by the behavioral state-specific gene content, we
260 annotated all OGs using eggnog and used COG categories as a first level of functional
261 characterization. Of all behavioral state-specific OGs, 33.1% had no COG annotation, while
262 17.3% of them fell into the COG category S (unknown function). For the rest of the significant
263 OGs, we categorized 20.6% of them as metabolism-related, 11.7% as linked to cellular
264 processes, signaling and structure, and 18.3% as information storage and processing **(Figure**
265 **3B**). Among the significant categorized OGs (excluding COG S), category G (carbohydrate
266 metabolism and transport) was the most important category, representing 11.4% of all
267 significant OGs. However, the proportion of different COG categories contributing to the
268 significant OGs varied between species **(Figure 3B**). In case of Firm4_1, COG category H
269 (coenzyme metabolism) was specifically enriched in the behavioral state-specific OGs relative
270 to the total number of OGs per species (Fisher's Exact test q-value=2.5e-06, **Supplementary**
271 **Table 4.**).

272 Behavioral state-specific OGs of COG category G included many phosphotransferase system
273 genes, in particular in Lactobacillus and Gilliamella species, but also glycoside hydrolases,
274 sugar binding and metabolism proteins. In most species, these OGs were more abundant in
275 foragers relative to nurses, except for Bifido_1.2 and Firm5_3, where the enrichment was in

276 the nurse samples, or spread equally between nurse and forager samples (**Supplementary**
277 **Dataset 1**).

278 The overrepresentation of the COG category H in Firm4_1 corresponded to 27 OGs involved
279 in the biosynthesis of cobalamin (vitamin B12) and adenosylcobalamin (coenzyme B12),
280 which had a higher coverage in forager than in nurse samples (see **Supplementary Dataset**
281 **1**). In the same species, we also found 18 genes for the utilization of ethanolamine among the
282 OGs enriched in foragers versus nurses. Processing of ethanolamine by the enzyme ammonia-
283 lyase is dependent on adenosylcobalamin (**Figure 3C**). To further explore the possible link
284 between the identified cobalamin biosynthesis and ethanolamine utilization OGs, we
285 analyzed the available isolate genomes of Firm4. Homologs of the cobalamin biosynthesis and
286 ethanol utilization pathways were only present in two (Hon2 and ESL0449) of the six Firm4_1
287 strains (**Figure 3D**) confirming that these genes are strain-specific. Moreover, the genes of
288 both pathways were encoded in the same genomic island, integrated into a conserved
289 chromosomal region, providing further evidence that they are functionally linked and have
290 likely evolved together.

291 Another interesting set of behavioral state-specific OGs was identified in Gilli_1, Gilli_2, and
292 Snod_1, and encoded various components of Type VI secretion systems (T6SSs) involved in
293 interbacterial killing. This included genes annotated as effector proteins with PAAR or RHS
294 domains and immunity proteins. In Gilli_2, several genes encoding the structural components
295 of an entire T6SS were among the behavioral state-specific OGs. They clustered in a genomic
296 island which was present in only one (strain A7) out of 17 strains of Gilli_2 in our reference
297 database (**Supplementary Figure 9**). Strikingly, all these behavioral state-specific T6SS-related
298 OGs were more abundant in forager than in nurse samples.

299 In Snod_1, we also found other surface structure-related functions than T6SS among the
300 behavioral state-specific OGs, including genes encoding major subunits of Type IV pili, RTX
301 toxins, and proteins involved in capsular polysaccharide biosynthesis (**Supplementary**
302 **Dataset 1**).

303 ## Discussion

304 There is accumulating evidence that natural microbial communities harbor high extent of
305 strain-level diversity [37-40]. How such fine-scale diversity varies across space and time in
306 nature is a key question in microbial ecology and evolution [15, 41, 42].

307 Here, we find that the strain composition of the honeybee gut microbiota undergoes a
308 consistent ecological turnover when bees transition from nurses to foragers. The single
309 nucleotide variant profiles of several community members clustered by behavioral state.
310 Moreover, for most community members, we found consistent differences in intraspecific
311 gene content between nurses and foragers, which included functions related to nutrient
312 utilization, vitamin biosynthesis, and cell-cell interactions.

313 In the human gut microbiota, strain-level variation in individual community members have
314 been linked to diet [43] or health state [42, 44-46]. Our results show that a radical shift in host
315 lifestyle can change the strain composition of a large fraction of the gut microbiota,
316 demonstrating that strains represent ecological distinct units and emphasizing the
317 importance to study diversity in host-microbe associations at the strain level.

318 Previous shotgun metagenomics studies have linked variation in species- and strain-level
319 diversity in the gut microbiota to host traits, such as age, colony origin, or genotype, but not
320 to behavioral state [22, 35, 47-49]. Some of these studies analyzed individual bees while
321 others focused on pooled gut samples. In the current work, we analyzed pooled samples of
322 20 bees rather than individual bees. The advantage of pooling is that a larger number of
323 colonies can be surveyed, increasing the statistical power of the study. However, pooling
324 precludes the detection of variants present at low prevalence across the sampled bees
325 explaining the overall lower levels of polymorphic sites observed in our as compared previous

326 studies [22, 35]. However, rare strains have likely a minor impact on the whole community or
327 the host.

328 Another possible limitation of our study is that the sampled nurse and forager bees were not
329 age-matched. As foragers are usually older than nurses, age could have been a confounding
330 factor in our study. However, our previous findings have shown that the age of bees has little
331 impact on the strain level composition of the gut microbiota [22]. Therefore, we are confident
332 that most of the observed differences are due to the nurse to forager transition and the
333 associated changes in dietary intake, physicochemical gut conditions, or host physiology.

334 In contrast to nurses, foragers do not consume pollen. This is likely the reason why foragers
335 were found to harbor lower bacterial biomass in the gut than nurses and exhibit a consistent
336 decrease of Bombilactobacillus, Lactobacillus, and Bifidobacterium relative to the other
337 phylotypes. This was already reported by two previous studies based on 16S rRNA gene
338 analysis [32, 33], one of which used age-matched forager and nurse bees, corroborating the
339 results of our shotgun metagenomics analyses. Pollen constitutes an important nutrient
340 source for bee gut-associated Bombilactobacillus, Lactobacillus, and Bifidobacterium
341 explaining their decrease in absolute abundance. All three phylotypes harbor genes to
342 degrade pollen-derived carbohydrates [21, 26, 50, 51], and they readily grow in minimal
343 media supplement with pollen extract or entire pollen grains [52, 53]. Moreover,
344 experimentally colonized bees kept on a sugar water diet harbor substantially fewer bacteria
345 of these phylotypes in their gut than bees that also receive pollen [33, 52].

346 As previous studies were based on 16S rRNA gene sequencing, their analyses were limited to
347 the phylotype level. Hence, a key finding of our study is that not only phylotypes, but also
348 closely related species within phylotypes, and even different strains of the same species

349 respond differently to the nurse-to-forager transition. For example, one species of
350 Lactobacillus Firm5 was similar in total abundance between foragers and nurses, while others
351 significantly differed significantly (>8 times). Likewise, some species showed consistent
352 changes in strain composition between nurses and foragers, while others did not. This
353 suggests that some species and strains are more successful than others in thriving in the gut
354 environment of one state versus the other. Reasons underlying these differences can be
355 manifold. Some species and strains may be adapted to utilize specific pollen compounds and
356 hence have an advantage in the nurse gut. Others may not depend on pollen nutrients for
357 growth that much; or may be better in withstanding physicochemical changes in the gut as
358 bees transition from one state to the other, explaining why they persist in the forager gut.
359 Also, the fact that we find changes in strain-level diversity in opposite directions between the
360 two host states suggests that different ecological processes are at play in different species.
361 For example, in one species, the reduced nutrient availability in the forager gut may hinder
362 strains to overgrow and dominate the community, resulting in decreased total abundance but
363 increased strain-level diversity in foragers. For another species, the nutrient-rich pollen diet
364 may expand ecological niches or open new ones that can be colonized by functionally
365 divergent strains explaining the observed increase in strain-level diversity in nurses relative
366 to foragers.

367

368 If the observed shifts in strain composition were due to changes in the ecological niches in
369 the gut, we would expect to find functional differences between strains of the same species
370 dominating in nurses versus foragers. Previous studies have successfully identified such
371 condition-specific genes within host-associated bacterial species by comparing metagenomic

372 datasets [17, 37, 43]. For example, strains of the human gut symbiont *Prevotella copri*
373 harbored distinct gene repertoires for drug metabolism and complex carbohydrate
374 degradation in Western and non-Western individuals [43]. Likewise, strains of an intracellular
375 symbiont of *Bathymodiolus* deep-sea mussels showed substantial variation in gene content
376 for nutrient and energy acquisition indicating adaptation to the local conditions present in
377 the environment they were sampled from [37].

378 We detected strain-specific gene sets (i.e. OGs) associated with one of the two behavioral
379 states in almost all species of the gut microbiota. Carbohydrate transport and metabolism
380 functions were prevalent among the behavioral state-specific gene content, which was
381 somehow expected as most species carry large accessory gene sets of this functional category
382 in their genomes [22, 50, 51]. Moreover, previous studies have shown that carbohydrate
383 transport and metabolism are important for niche partitioning of diet-derived carbohydrates
384 in Lactobacillus and Bifidobacterium and that many of these genes are differentially
385 expressed when bees are fed pollen as compared to sugar water as diet [26, 52].

386 Somewhat more surprising was the finding that vitamin B12 biosynthesis and ethanolamine
387 utilization genes of Bombilactobacillus were specifically enriched in foragers versus nurses.
388 Some of these genes have already been found to be variably present among individual bees
389 in our previous study [22], but they could not be linked to any host differences. Vitamin B12
390 is an important co-factor that cannot be synthesized by honeybees (or other animals) [54],
391 but needs to be taken up from the diet or synthesized by bacterial symbionts [55]. It is
392 possible that Bombilactobacillus strains synthesizing this vitamin compensate the lack of
393 vitamin B12 in the pollen-deprived diet of foragers. However, a more plausible explanation
394 for the increased abundance of these genes in foragers may be that vitamin B12 is an essential

395 co-factor for a major enzyme of the ethanolamine utilization pathway [54]. Ethanolamine is
396 derived from the membrane phospholipid phosphatidylethanolamine and constitutes an
397 important bacterial nitrogen and carbon source in the animal gut due to the high turnover of
398 the host epithelium and the microbiota [56]. It is possible that ethanolamine accumulates in
399 the forager gut due to the presence of dead bacterial material or increased epithelial
400 shedding, providing a behavioral state-specific nutrient niche for the gut microbiota. The fact
401 that the genes for utilizing ethanolamine are strain-specific indicates that they are costly to
402 maintain, and that host state-specific nutritional niches may contribute to the maintenance
403 of strain-level diversity in the gut microbiota.

404 We also found genes mediating bacteria-host interactions and interbacterial killing among
405 the behavioral state specific gene content. Of particular interest was the finding of several
406 T6SS genes associated with foragers in two of the three Gilliamella species and in
407 Snodgrassella. The primary niche of these bacteria is the ileum, where they form multispecies
408 biofilms on the host epithelium [57]. Diverse T6SS gene repertoires have been identified in
409 almost all sequenced strains of Gilliamella and Snodgrassella indicating that they play key
410 roles in interbacterial competition and niche establishment for these bacteria [58, 59]. Our
411 findings suggest that the forager gut environment selects for strains with expanded
412 capabilities in interbacterial warfare, possibly as a consequence of the increased competition
413 for nutrients and space. In Snodgrassella, we also found genes encoding other surface
414 structures (Type IV pili, LPS modifying enzymes, RTX toxins) to be enriched in either nurses or
415 foragers. Some of these genes are known colonization factors of Snodgrassella [60], likely
416 because they enable host attachment and biofilm formation. However, why these functions
417 are specifically enriched in in one of the two behavioral states remains to be determined.

418

419 ## Conclusions

420 In summary, using shotgun metagenomics analysis we show that the gut microbiota of
421 honeybees undergoes consistent shifts in strain-level diversity when bees transition between
422 two major behavioral states (i.e. from nurses to foragers). These strain-level differences are
423 associated with changes in intraspecific gene content of several bacterial species suggesting
424 adaptation to host state-specific niches. The possibility to carry out gnotobiotic bee
425 experiments provides excellent opportunities for future research to probe the functional
426 relevance of the link between strain-specific gene content and host state. Since the
427 association of social bees with their microbiota is evolutionary ancient, it will also be
428 interesting to expand these analyses to closely related bee species which show similar division
429 of labor and whose microbiota has diversified to similar extents as in *A. mellifera*.

430

431 ## Material and methods

432 ### Sampling

433 Adult female worker bees of *A. mellifera* were collected over the Summer 2019 from 15
434 different colonies in 5 different locations across Western Switzerland (**Supplementary Table**
435 **1**). For each sampled hive, we collected 20 larvae-feeding nurses and 20 pollen foragers, in
436 order to encompass most of the strain-level diversity within each colony. Samplings were
437 performed between 7 AM and 9 AM, when worker bees were beginning to return from their
438 foraging trips. Larvae-feeding nurses were identified as bees putting their heads in larvae
439 containing cells. Pollen foragers were identified as bees returning to their hive with pollen
440 pellets on their corbiculae. The bees were anesthetized using CO_2 and put on ice. They were
441 then dissected in sterile 1X PBS to extract the entire hindgut (pylorus, ileum, rectum) with
442 tools sterilized in 70% ethanol. Gut content was visually inspected to be consistent with life-
443 stage, i.e. nurse guts had to be pollen filled and forager guts had to be nectar-filled. Bees for
444 which the gut content could not be verified were excluded. Finally, the guts were pooled by
445 hive of origin and host phenotype (nurse and forager, respectively) (**Supplementary Table 1**).

446

447 ### DNA extraction

448 DNA extraction from the pooled gut samples was performed using the Qiagen Allprep
449 Powerfecal DNA/RNA extraction kit (Qiagen) following the kit handbook except for the bead-
450 beating steps. After dissection, the pooled guts were transferred into the kit's bead-beating
451 tubes. Lysis buffer and DDT were added following protocol instructions, and samples
452 underwent two bead-beating steps of 4 seconds at 6 m/s on a FastPrep-24 5G bead-beater
453 (MP Biomedicals). Gut samples with total weight over 220 mg were diluted to reach a

concentration of 220 mg in 650 µl lysis buffer + 25 µl DDT, then excess volume was removed (i.e. for a gut weight of 440 mg, 337.5 µl were removed from the sample, and an identical volume of lysis buffer + 3.8% DDT 1M was added to it). Subsequently, all samples underwent a third bead-beating step for 45 seconds at 6 m/s. All the remaining steps were performed following the manufacturer's protocol. DNA was diluted in 2x30 µl Buffer EB, its concentration and quality were assessed with Qubit and Nanodrop, and samples stored at -20°C until further processing.

Quantitative real-time PCR

We assessed the bacterial loads of the DNA samples using quantitative real-time PCR (qPCR) following the methods of Ellegaard et al. (2020). We targeted both the V3-V4 region of the 16S rRNA bacterial gene and the actin gene of *A. mellifera* using the primers published by Kešnerová et al (2017). We performed each qPCR reaction using a reaction of 1 µl extracted DNA (10 ng/ µl), 3.2 µl nuclease-free water, 5 µl SYBR Select Master Mix (Life Technologies), 0.2 µl forward primers (5 µM) and 0.2 µl reverse primers (5 µM) in 96-well plates (MicroAmp Fast 96-well Reaction Plate, Applied Biosystems for Life Technologies). The reactions were performed in triplicates on a QuantStudio 5 Real-Time PCR System (Thermo Fischer Scientific) using the parameters from Kešnerová et al (2017).

The data was extracted using the QuantStudio Design & Analysis Software v1.4 (Applied Biosystems) and exported in csv format.

We computed the absolute number of each gene copy using the formula:

$$gene\ copy\ number = 10^{\frac{C_t - Intercept}{Slope}}$$

476 Intercept and slope parameters previously determined by our group [53].

477 In order to compare the microbial gut loads of the different samples, we normalized the
478 number of 16S rRNA gene copies by the total DNA yield per sample. Normalization by the
479 number of host actin gene copies gave similar results. However, foragers had more actin gene
480 copies than nurses. Therefore, we did not use this type of normalization in the final analysis
481 (see **Supplementary Figure 10**). The figures were generated using R (version 4.0) and the
482 ggplot2 library.

483

484 **Shotgun metagenomics sequencing**

485 The libraries were prepared with the Illumina Nextera Flex library kit (Illumina) with unique
486 dual indices (UDI) and sequenced on a HiSeq 4000 instrument (PE150) at the Genomic
487 Technologies Facility (GTF) of the University of Lausanne. All samples were multiplexed
488 together and sequenced on three lanes. The quality of the reads was assessed with FastQC
489 (v0.11.4, Babraham Institute) and subsequently trimmed and filtered to remove low quality
490 sequences and short reads using the tool Trimmomatic v0.35 [61] with settings PE
491 ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50.

492

493 **Reference genome database**

494 The reference genome database containing bacterial isolate genomes was constructed in the
495 same way as in Ellegaard and Engel (2019) and Ellegaard et al. (2020). A total of 198 high-
496 quality genomes of strains isolated from the gut of honeybees and other social bee species
497 such as bumble bees and other *Apis* species were included in the database. Strains having

498 more than 98.5% pairwise ANI with other already included strains were not added to the
499 database to limit redundancy. The pairwise genomic ANI was computed using fastANI [62]. In
500 general, multiple genomes were included for each bacterial phylotype as listed in
501 **Supplementary Table 2**. Each genome includes a fasta file of nucleotide sequences of all
502 contigs (.fna), nucleotide sequences (.ffn) and amino acid sequences (.faa) of predicted genes
503 and concatenated (.fna) files where the contigs were concatenated into one sequence for
504 each genome. Bed files were made to summarize the updated positions of genes in the
505 concatenated files which were used for further analysis. We predicted gene families of
506 homologous genes using OrthoFinder (v2.3.5, [63]), and identified all single-copy core genes,
507 i.e. genes present in one copy in all genomes of a given phylotype. The resulting files listing
508 each single-copy core orthologous group and the constituent genes of each genome of that
509 phylotype are included in the reference genome database along with reduced bed files for
510 each genome indicating the location of these core genes to be used for further analysis. The
511 total length of these single-copy core genes relative to the total genome length of a given
512 species are given in **Supplementary Figure 11**.

513

514 **Community composition analysis**

515 The community composition analysis was based on the work of Ellegaard and Engel (2019).
516 We mapped the reads using bwa mem (v0.7.15-r1142-dirty) [64] against the reference
517 genome database and used the samtools (v1.9) view –F 0x800 command [65] to remove
518 chimeric alignments. Reads mapping to multiple locations were not included as bwa chooses
519 one of the many alignments as primary and only includes those in the output. Alignments
520 with 50 matches or fewer were removed using a custom perl script that parses alignment

521 length from the CIGAR string. The read coverage information was extracted for each gene
522 using the command samtools bedcov and summed up across all orthologs of each single-copy
523 core gene family for genomes coming from the same species. Since genes located close to the
524 origin of replication had higher coverage than genes located close to the terminus (as an
525 indication of replicating bacteria [66]), the coverage at the terminus was inferred. To this end,
526 the coverage of each gene family was plotted relative to its genomic position using one
527 selected reference genome per species and a segmented regression line was fitted. The
528 relative abundance of each species in each sample was determined by dividing the terminus
529 coverage by the summed terminus coverage of all species, respectively. For the species
530 Bifido_1.1, Bifido_1.2 and Bifido_1.4, the segmented lines did not show a proper fit (terminus
531 location inferred far from estimated breakpoint, or coverage at *Ori* lower than *Ter*). In these
532 cases, the median coverage of all single-copy core gene families was used as an estimate for
533 the relative abundance of that species.

534

535 **Strain-level analysis**

536 To quantify strain-level diversity within species we followed the approach established in
537 Ellegaard and Engel (2019). In short, we mapped the metagenomic reads using bwa mem to
538 a reduced version of the reference genome database, including only one representative
539 member per species (see **Supplementary Table 2**). Then, we used samtools -F 0x800 to
540 remove chimeric alignments. Reads mapped with less than 50 matches registered in their
541 CIGAR were also removed, as well as those whose number of mismatches was higher than 5
542 to reduce errors in variant calling. Only species with more than 20x terminus coverage were
543 included. Genes with less than 10x coverage as estimated in the community profiling pipeline

544 were excluded. To perform variant calling, we used the tool FreeBayes (v1.3.1-19-g54bf409)
545 [67] with options -—haplotype-length 0, --min-alternate-fraction 0.1, —-min-coverage 10, -C 5,
546 -I, -X, -u and —-pooled-continuous. We used the tools vcfbreakmulti and vcfallelicprimitives
547 from the vcflib library [68] to split complex observations into multiple lines. The bed files
548 containing gene positions of single-copy core genes of each genome were subset to obtain
549 the corresponding locations to be considered for each species. These filtered bed files were
550 used to focus downstream analysis to the phylotype-core region for each species.

551 We computed the fraction of polymorphic sites for each bacterial species using the number
552 of identified sites divided by the cumulative length of all single-copy core genes. An allele was
553 considered detected if there was more than one read supporting the allele and the allele was
554 at a frequency of at least 0.01. A position was considered polymorphic if more than one allele
555 was detected within the metagenomic reads of that sample mapped to the reference
556 regardless of what allele was in the reference genome. We also computed the cumulative
557 number of polymorphic sites across samples for each species using 10 random orders of
558 samples each time.

559 To assess the similarity between samples in terms of strain composition, we used the Jaccard
560 index based on the proportion of shared polymorphic sites; i.e. sites which were polymorphic
561 in at least one of the samples in each pair and where each sample had an allele and an
562 alternate allele shared and at a frequency above the detection threshold (0.01) were
563 considered shared between the two samples. So, for example, if the reference genome has
564 the allele T at a given position, but sample 1 has the alleles C and G and sample 2 has the
565 alleles T and C, then the two samples would share one allele (C) and they would have two
566 non-shared alleles. This site would not be counted as shared. On the other hand, if sample 1

567 had the alleles C and T the polymorphic site would be considered shared with sample 2. This
568 was implemented in a custom python script as cases where the site was polymorphic in at
569 least one sample in the pair were scored and shared positions were those with an intersection
570 > 1 in the set of detected alleles in the sample pair. The distance between all sample pairs
571 was computed as $Distance(S1, S2) = 1 - Shared\ fraction(S1, S2)$ where shared fraction
572 for a given pair of samples was calculated as the number of shared polymorphic sites divided
573 by the number of sites polymorphic in at least one of the two samples. The matrix of pairwise
574 distances between the samples was plotted using principal coordinate analysis to identify
575 samples that share similar polymorphic sites. We computed the principal coordinates using
576 the ape R library (v5.6-2).

577

578 **Functional gene content composition**

579 For the functional profiling of the gene content of the microbiota, we first extracted all the
580 reads which mapped to the bacterial database, plus those which did not map to the host
581 genome (*A. mellifera* genome (PRJNA471592, version Amel_Hav3.1), using the picard tools
582 (v2.7.1-SNAPSHOT) SamToFastq with parameters VALIDATION_STRINGENCY=SILENT [69]. We
583 merged those reads for each sample and assembled them using SPAdes [70] with parameters
584 --meta and -t 20. The obtained contigs were filtered to keep only the ones with a length of at
585 least 500 bp and a k-mer coverage of at least 1.

586 We predicted the open-reading frame (ORF) positions using Prodigal (version 2.6.3, Hyatt et
587 al., 2010) [71] with parameters -m -p meta. The ORFs with the flag partial=00 or length smaller
588 than 300 bp were filtered out. We annotated the remaining ORFs using eggNOG (v1.0.3-40-

589 g41a8498) [72, 73] with the script emapper.py and the options --annotate_hits_table --
590 no_file_comments --cpu 20.

591 To assign the contigs to the major bacterial species of the bee gut microbiota, we blasted the
592 predicted ORFs against a nucleotide database of all ORFs from our bacterial database using
593 the blastn tool (version 2.2.31+) [74] (evalue < 1e-05, pident > 50 and qcovs > 50). Contigs
594 were assigned to a given species when at least 80% of the ORFs were mapped against genes
595 of that species, as well as when at least 10% of the ORFs were mapped against genes of that
596 species and no mapping was found for any other species. We filtered out all non-unanimous
597 contigs (15,198 contigs out of a total of 257,015 (5.9%), representing 13.6% of the bp).

598 For each species, we used the amino acid sequences from the ORFs and the genes from the
599 database strains to compute orthologous groups (OG) using OrthoFinder [75]. We created a
600 database of all ORF's nucleotide sequences and mapped the reads against it using bwa mem
601 (version 0.7.15-r1142-dirty) [64]. We removed chimeric alignments using samtools view
602 option –F 0x800 (version 1.9)[65]. Alignments with less than 50 matches were also removed.
603 For each sample, we extracted the coverage using the samtools coverage command (v1.15),
604 and summed the coverage for the ORFs belonging to the same OG. We normalized the
605 coverage information for every sample and every species using the OGs containing single-
606 copy core genes: for each species of each sample, the summed coverage of all single-copy
607 core OGs was set to the same number (10000). To determine which OGs were
608 underrepresented or overrepresented in either the nurses or the foragers, we computed the
609 paired Wilcoxon signed rank test for each OG based on the normalized coverage values, and
610 then used q-values to correct for multiple comparisons. We also computed the fold-change
611 by dividing the mean forager normalized coverage value by the nurse counterpart. To be

612 considered as significantly different between the treatments, an ORF had to have both an
613 absolute log2 fold change of at least 1, and a q-value < 0.05. We finally annotated the ORFs
614 belonging to the significant OGs using the tool eggnog-mapper (1.0.3-40-g41a8498)[72, 73].
615 The whole analysis workflow is summarized in **Supplementary Figure 8**.

616

617 **Statistics and Data visualization**

618 We performed the statistical analyses in R (v4.2.1) using the libraries ape (v5.6-2), data.table
619 (v1.14.8), and vegan (v2.6-4) and in python (v3.10.11) using the package scipy (v1.11.1). We
620 used the package qvalue (v2.28.0) to adjust p-values to q-values where stated. Shannon
621 indices were calculated using the vegan package. For PERMANOVA we used the function
622 adonis2 from the vegan package using the model "Distance ~ Host + Location" with 999
623 permutations and set to estimate the marginal effects of the terms. R2 values were adjusted
624 using the function adonisOmegaSq implemented in the MicEco package
625 (https://github.com/Russel88/MicEco, commit 91f8e6f).

626 We generated all plots using R and the following R libraries: ggplot2 (v3.3.6), reshape (v0.8.9),
627 gridExtra (v2.3), ggrepel (v0.9.1), RColorBrewer (v1.1-3) and scales (v1.2.1) and in python
628 (v3.10.11) using the package matplotlib (v3.7.1) and seaborn (v0.12.2). Figures have been
629 generated using Inkscape (v1.1) or Adobe Illustrator. A comprehensive list of all packages
630 used and their version is available in the github repository.

Data availability

The sequencing data is available under the BioProject ID PRJNA904667. The honeybee gut bacterial strain genome database is available at https://drive.switch.ch/index.php/s/vHmM6aIIFyDGQwm. The scripts required to run the analyses and produce the figures are available on GitHub (https://github.com/Aiswarya-prasad/Publication_Baud_metaG_NvsF_2022).

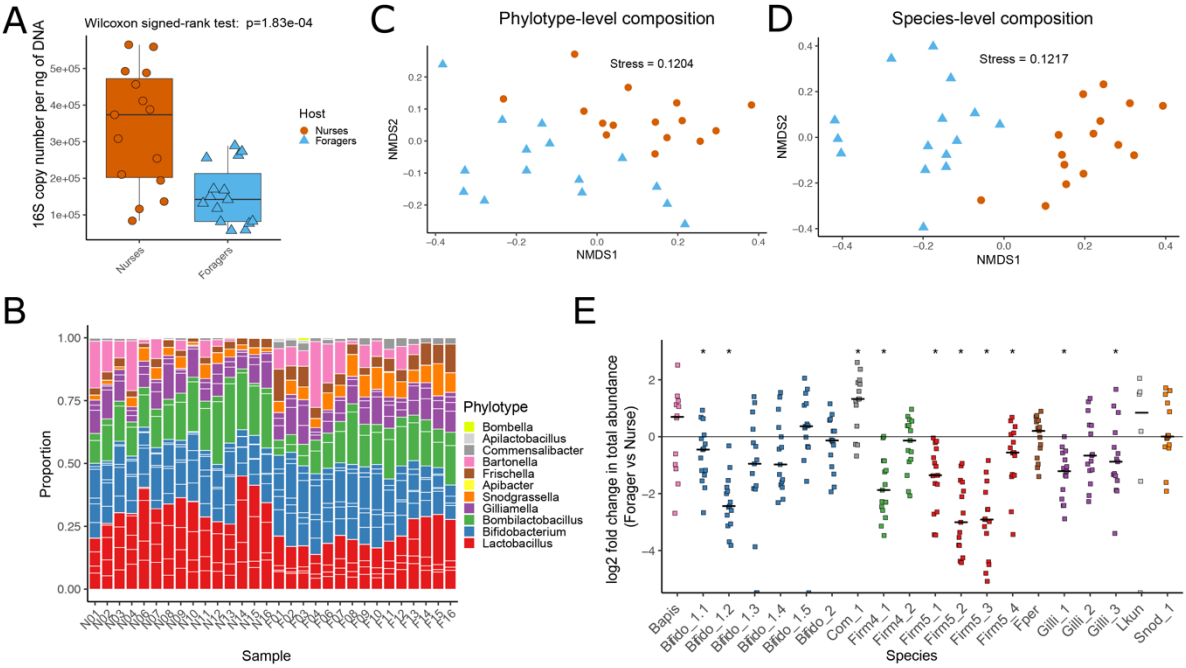
Acknowledgements

We would like to thank Mam Malick Sy Ndiaye, Florent Mazel, Garance Sarton-Lohéac, Shinichi Sunagawa for providing helpful comments on the manuscript. We would like to thank Julien Marquis and the team from the Lausanne Genomics Technology Facility for carrying out the Illumina sequencing. We would like to thank Benjamin Dainat, Jan Dudda, Olivier Emery and Chloé Peillon for letting us sample bees from their apiaries. This work was funded by the University of Lausanne, an ERC Starting Grant (MicroBeeOme), the NCCR Microbiomes, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant no. 180575) and a Swiss National Science Foundation project grant (grant no. 179487) to P.E.

Conflict of interest statement

None of the authors has any competing interests.

652 **Figure legends**



653

654 **Figure 1.** Community composition analysis of the gut microbiota of nurse and forager bees.
655 (A) Boxplot of the bacterial loads of each gut represented by the number of bacterial 16S
656 rRNA gene copies relative to 1ng of total DNA. (B) Stacked bar plot of the bacterial phylotype
657 composition for each sample. White separators within a phylotype delineate the different
658 species. (C) NMDS plot of the samples' bacterial phylotype composition. (D) NMDS plot of the
659 samples' bacterial species composition. (E) Log2 fold change in total normalized absolute
660 abundance of bacterial species in forager samples relative to nurse samples from the same
661 hive. Asterisks mark species with significant differences between nurses and foragers
662 (Wilcoxon signed-rank test of the normalized absolute abundance of each species in the
663 foragers against the nurses, q-value < 0.05). Data points for species of the same phylotype
664 are shown in the same color. Horizontal bars represent medians.

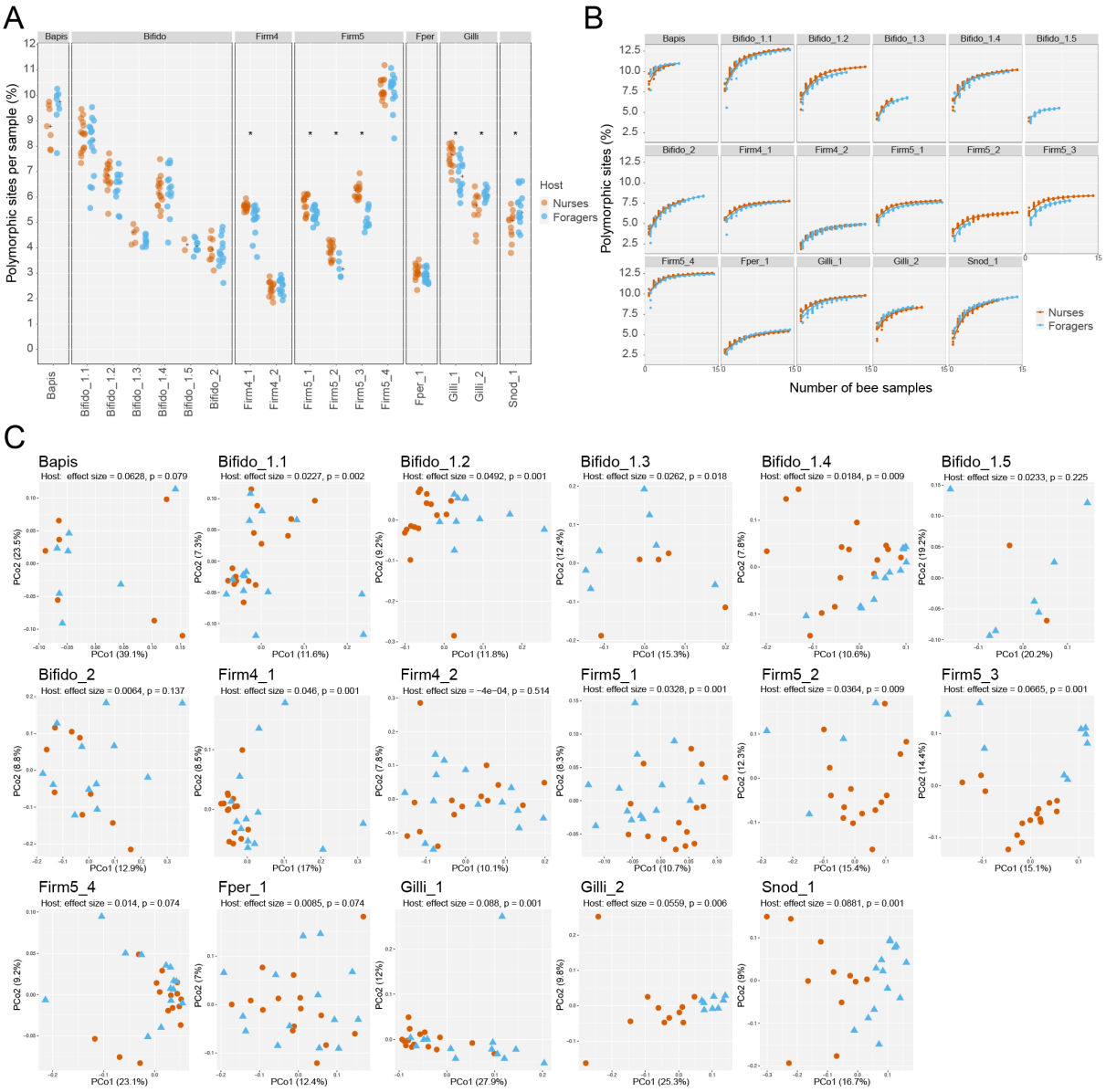


Figure 2. Differences in strain-level diversity between nurse and forager bees. (A) Proportion of polymorphic sites in the single-copy core genes for each bacterial species, expressed as a percentage of the total single-copy core genome length. Asterisks symbolize a q-value < 0.05 for the Wilcoxon signed-rank test between the proportion of polymorphic sites between nurse and forager samples. (B) Curves of the cumulative proportions of polymorphic sites for each species. (C) Principal Coordinate Analysis plots based on the Jaccard distance shared polymorphic sites between each pair of samples for each of the species.

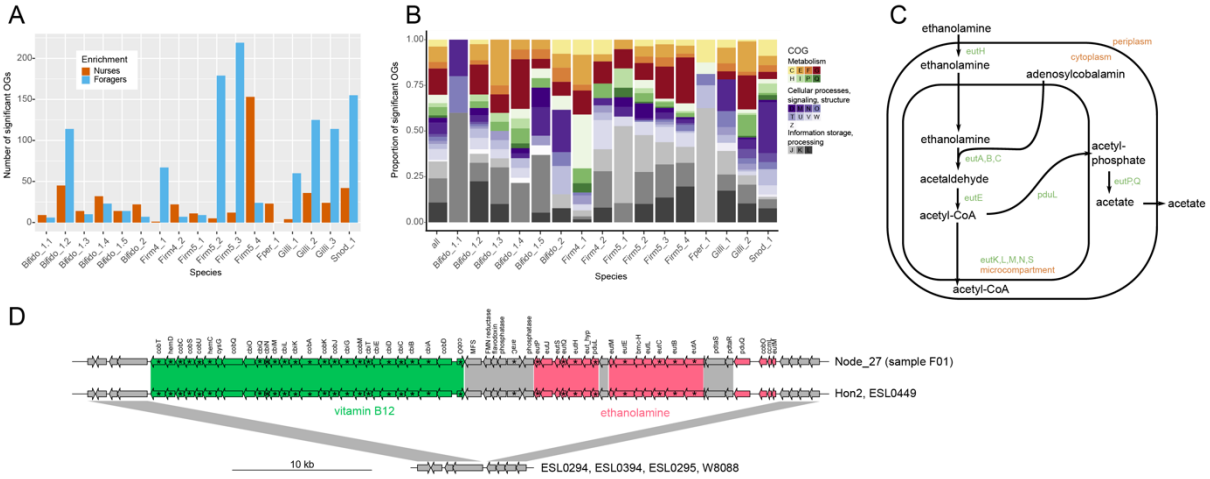
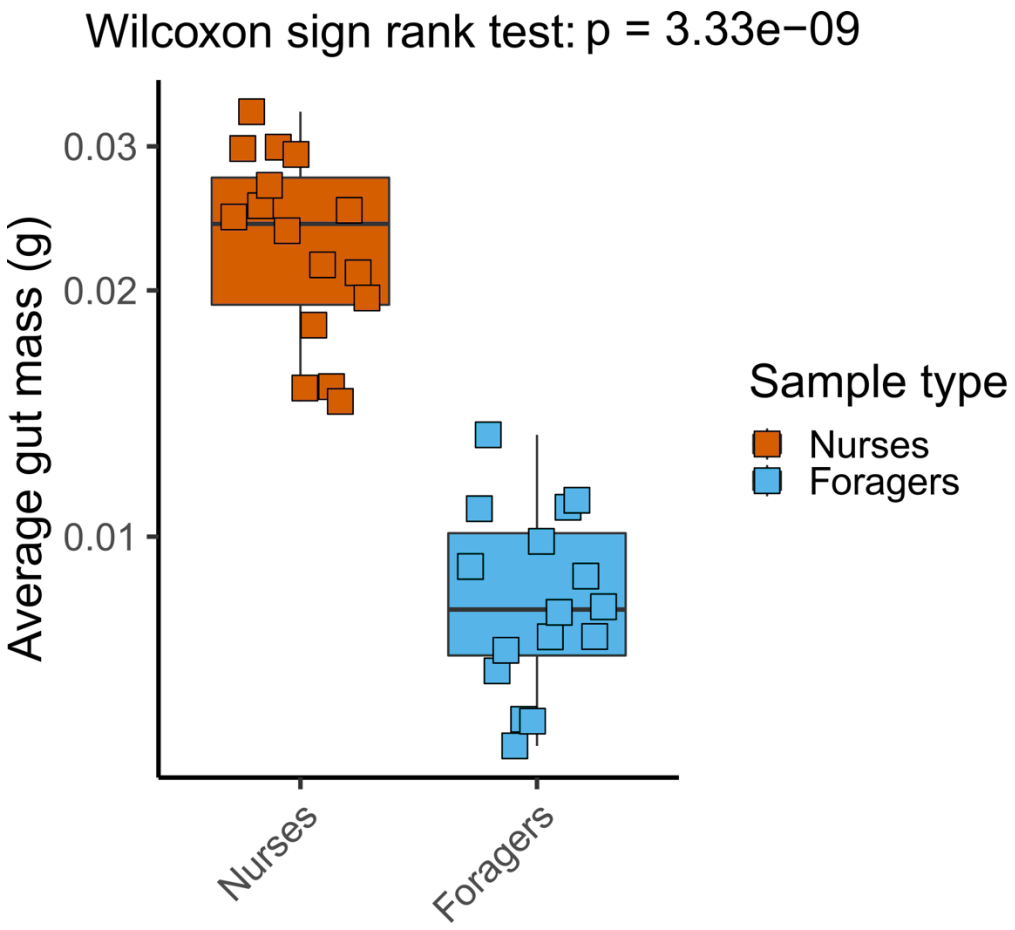


Figure 3. Differences in functional gene content within species of the gut microbiota between nurse and forager bees. (A) Number of orthologous groups (OGs) significantly differing in coverage for each of the species. (B) Repartition of the COG categories of the significant OGs, nurse-enriched and forager-enriched included. OGs that were not categorized or categorized as COG S were not included in the plot. (C) Figure adapted from [76]. Possible ethanolamine utilization pathway enabled by the operon shown in (D). (D) Genomic island containing the ethanolamine utilization and vitamin B12 biosynthesis genes, and its presence in the isolate genomes of the bacterial genome database. Genes enriched in the forager samples are in bold and flagged with an asterisk.

685 **Supplementary tables and figures**

686

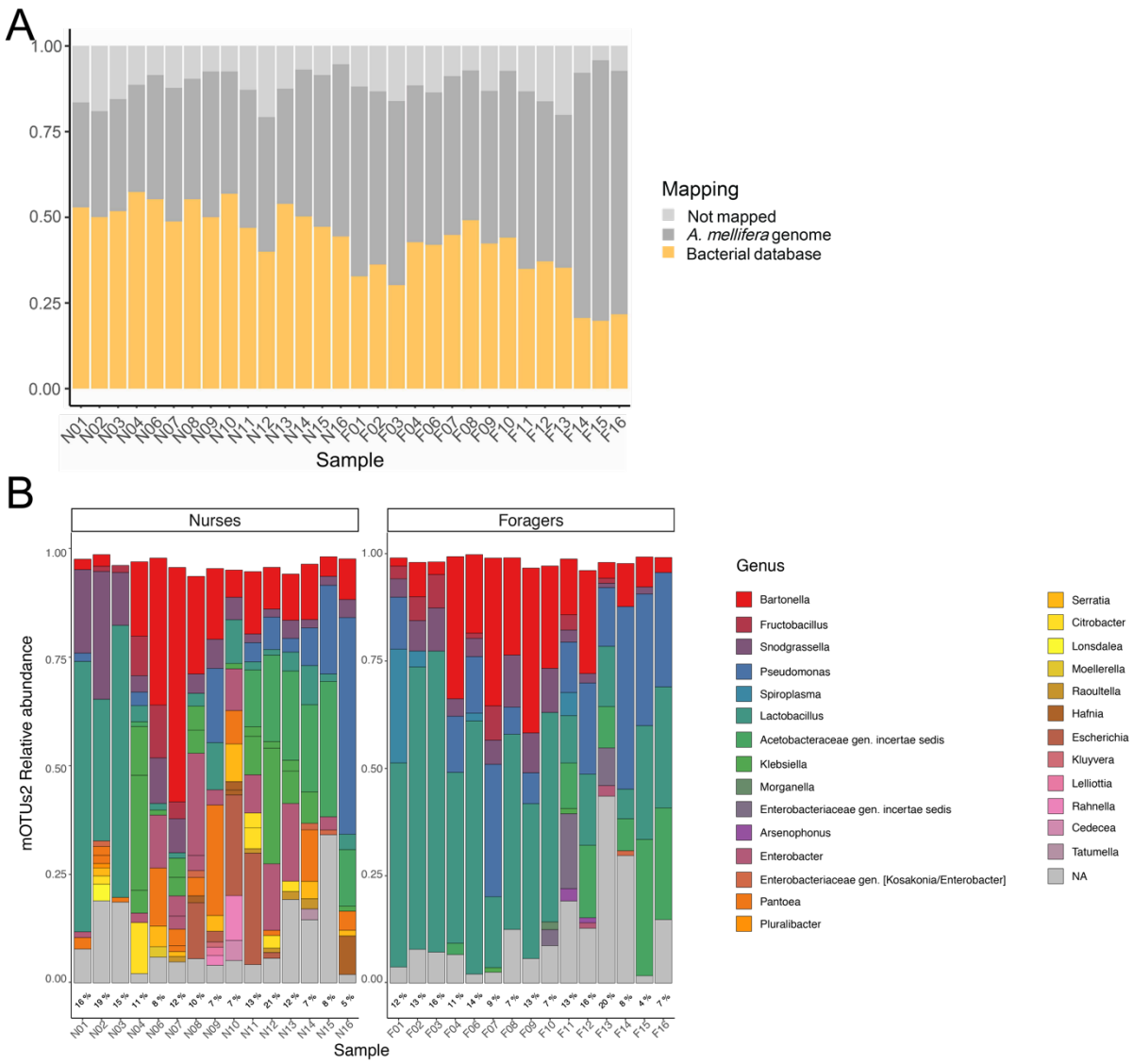


687

688 **Supplementary Figure 1.** Per sample average gut mass, computed as each sample's gut pool

689 mass, divided by the number of guts in the pool.

690



691

692 **Supplementary Figure 2.** (A) Stacked bar plot of the proportions of reads mapping to the

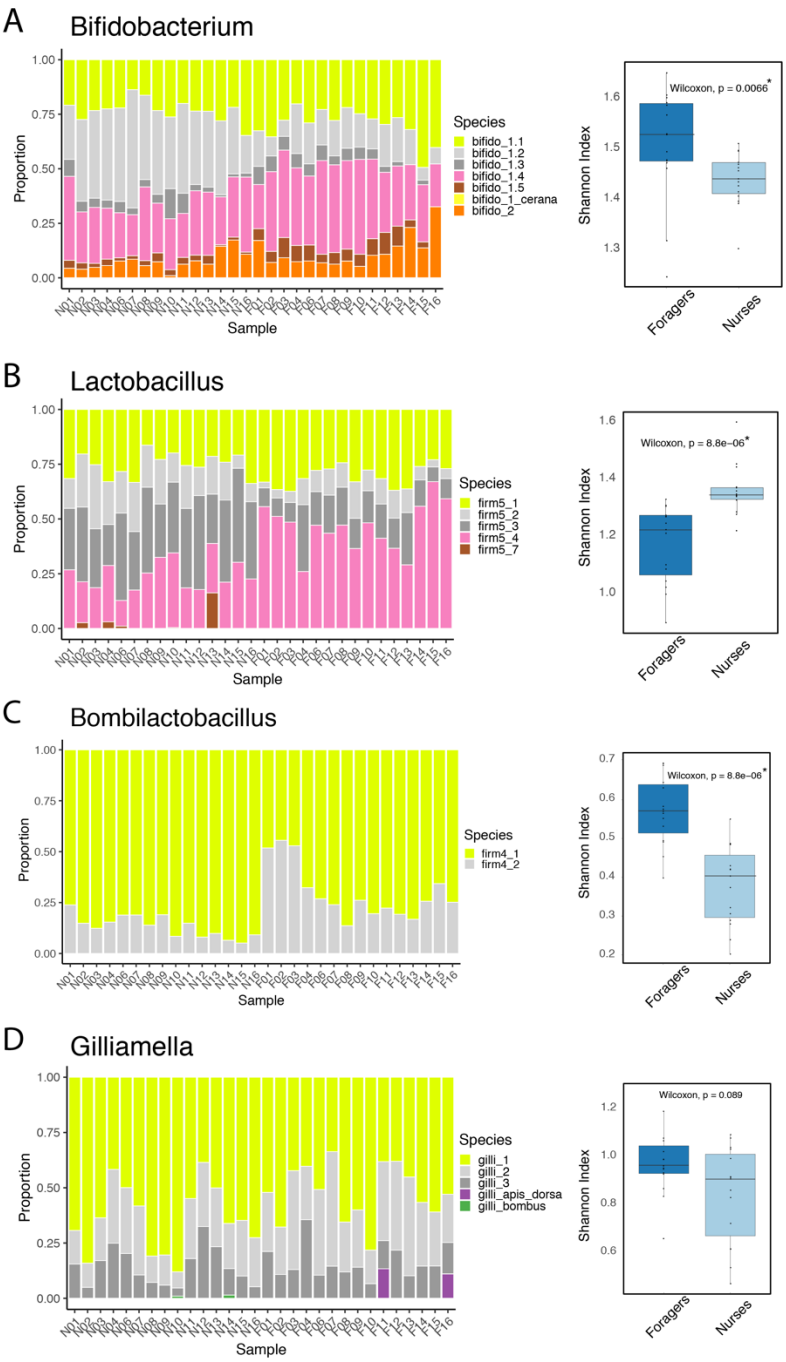
693 bacterial genome database and to the *A. mellifera* genome, as well as non-mapped reads, for

694 each of the samples. (B) Taxonomic profiling of the unmapped reads at the species-level with

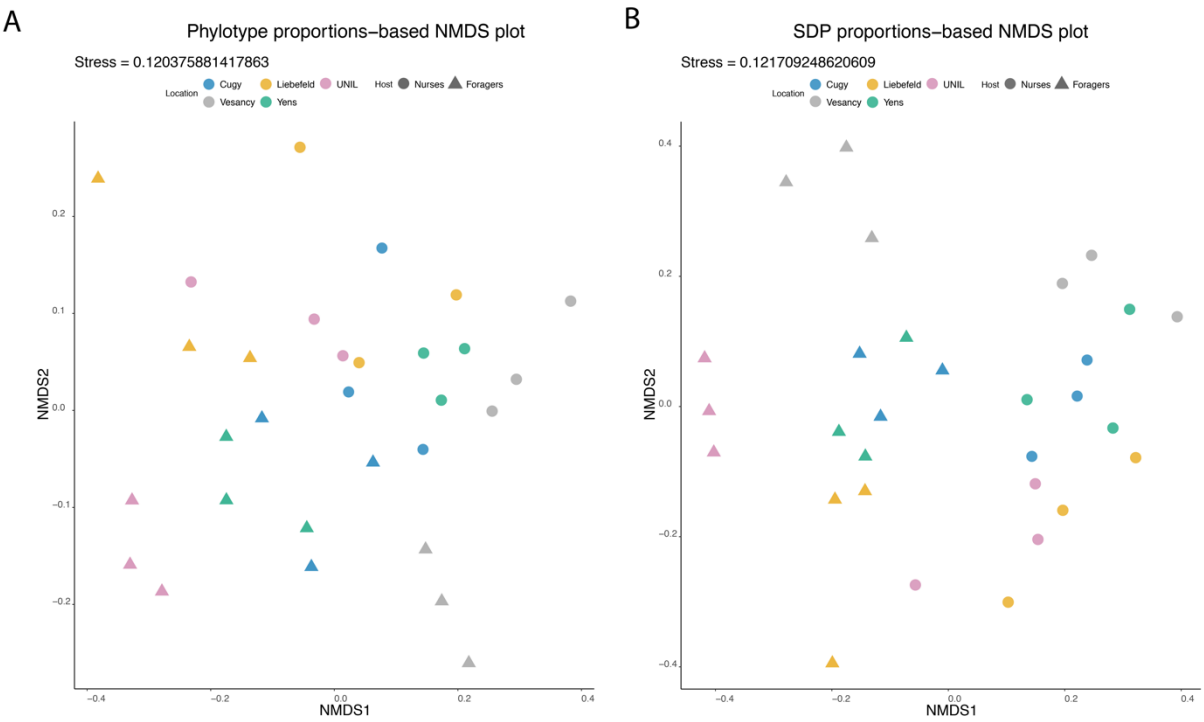
695 mOTU2. Colors correspond to genera. Percentages below each distribution show fraction of

696 unmapped reads in each sample (also see panel A).

697



Supplementary Figure 3. Stacked bar plots of the proportions of the different species and corresponding Shannon diversity index across nurse and forager samples for the four multi-species phylotypes (A) Bifidobacterium, (B) Lactobacillus, (C) Bombilactobacillus, and (D) Gilliamella.

704 

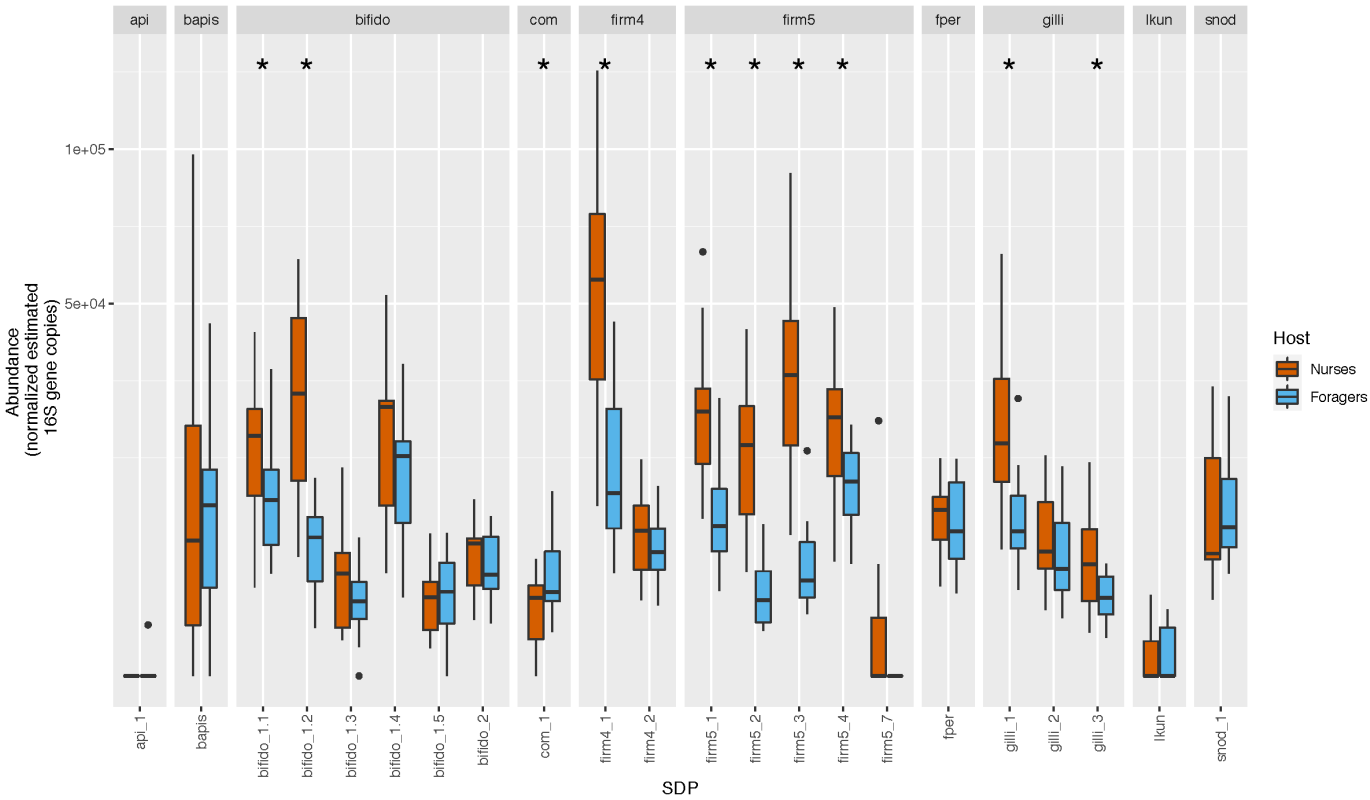
705 **Supplementary Figure 3.** NMDS plot of the samples' bacterial phylotype composition (A)

706 and species composition (B). Samples are colored according to sampling location.

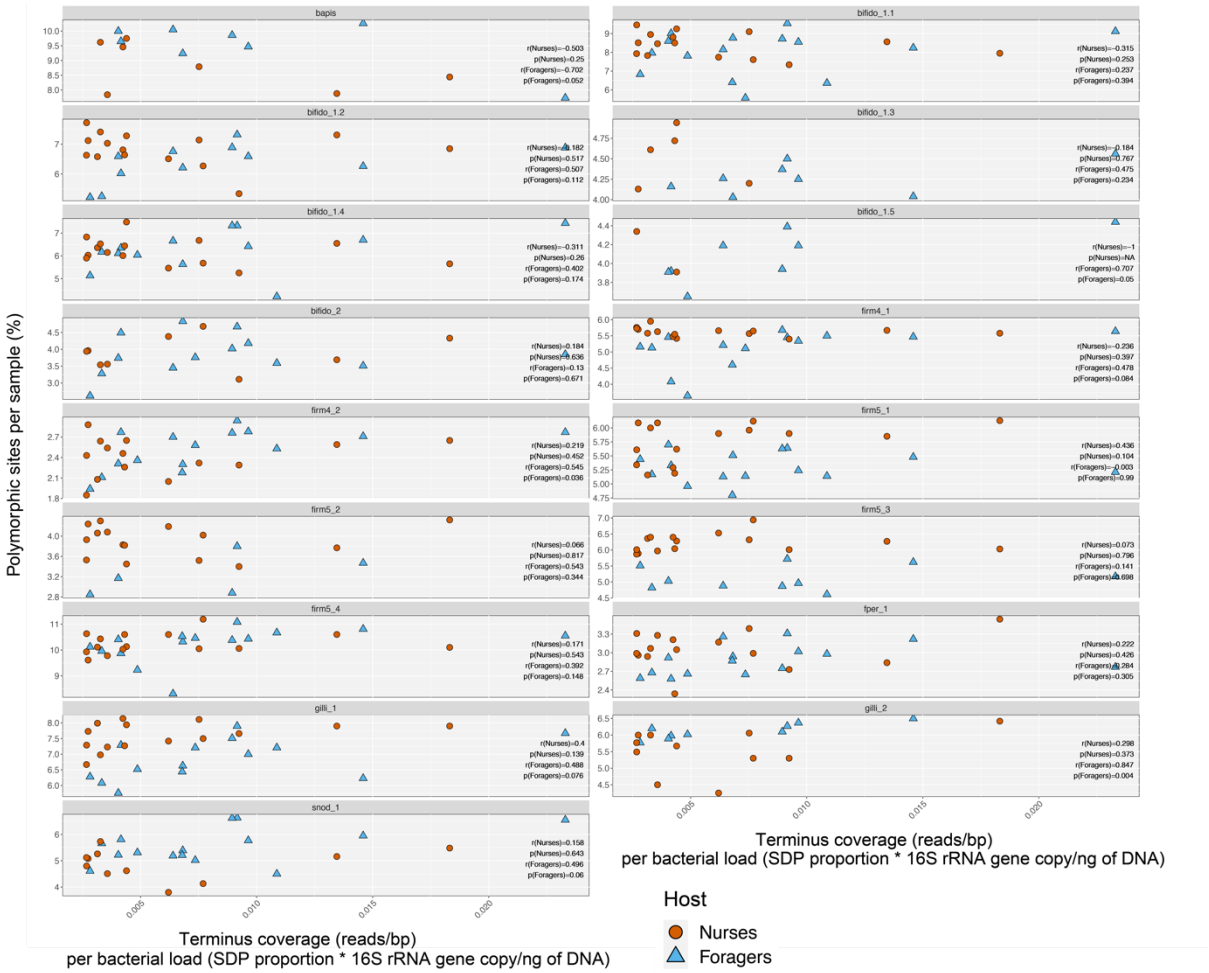
707 PERMANOVA results: Phylotype-level, Location – R^2=0.40279, p-value=0.001; Species-level,

708 Location – R^2=0. 40642, p-value=0.001.

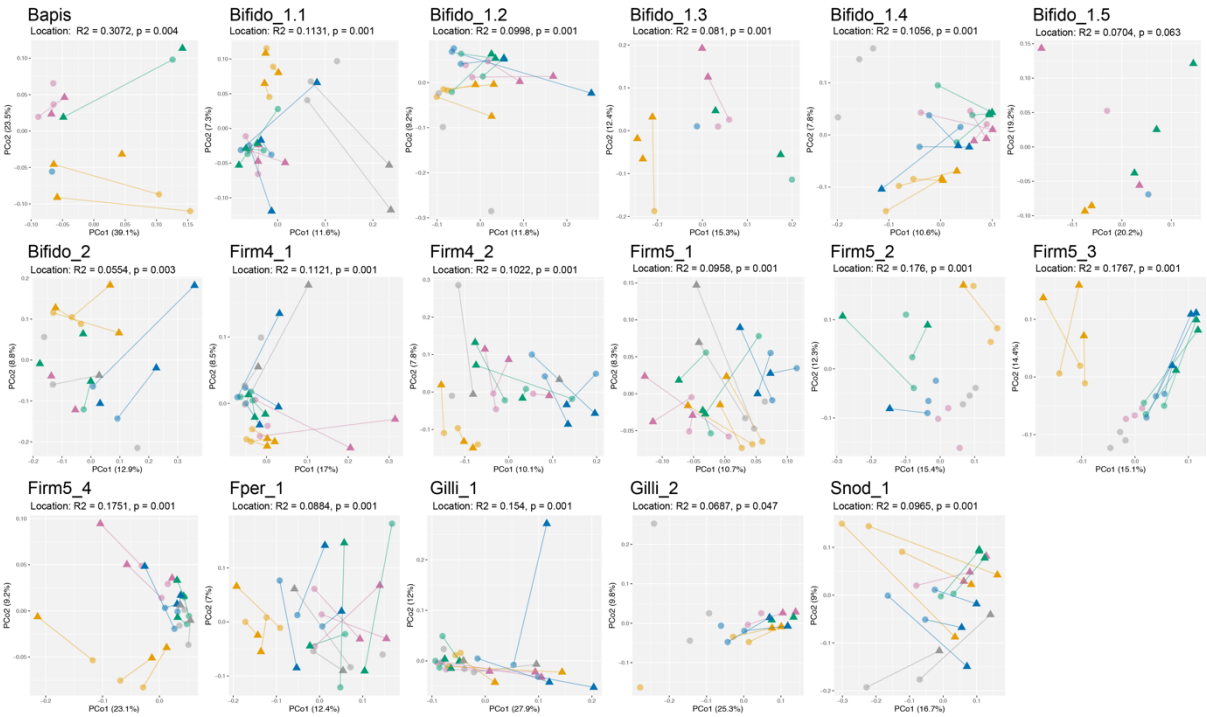
709



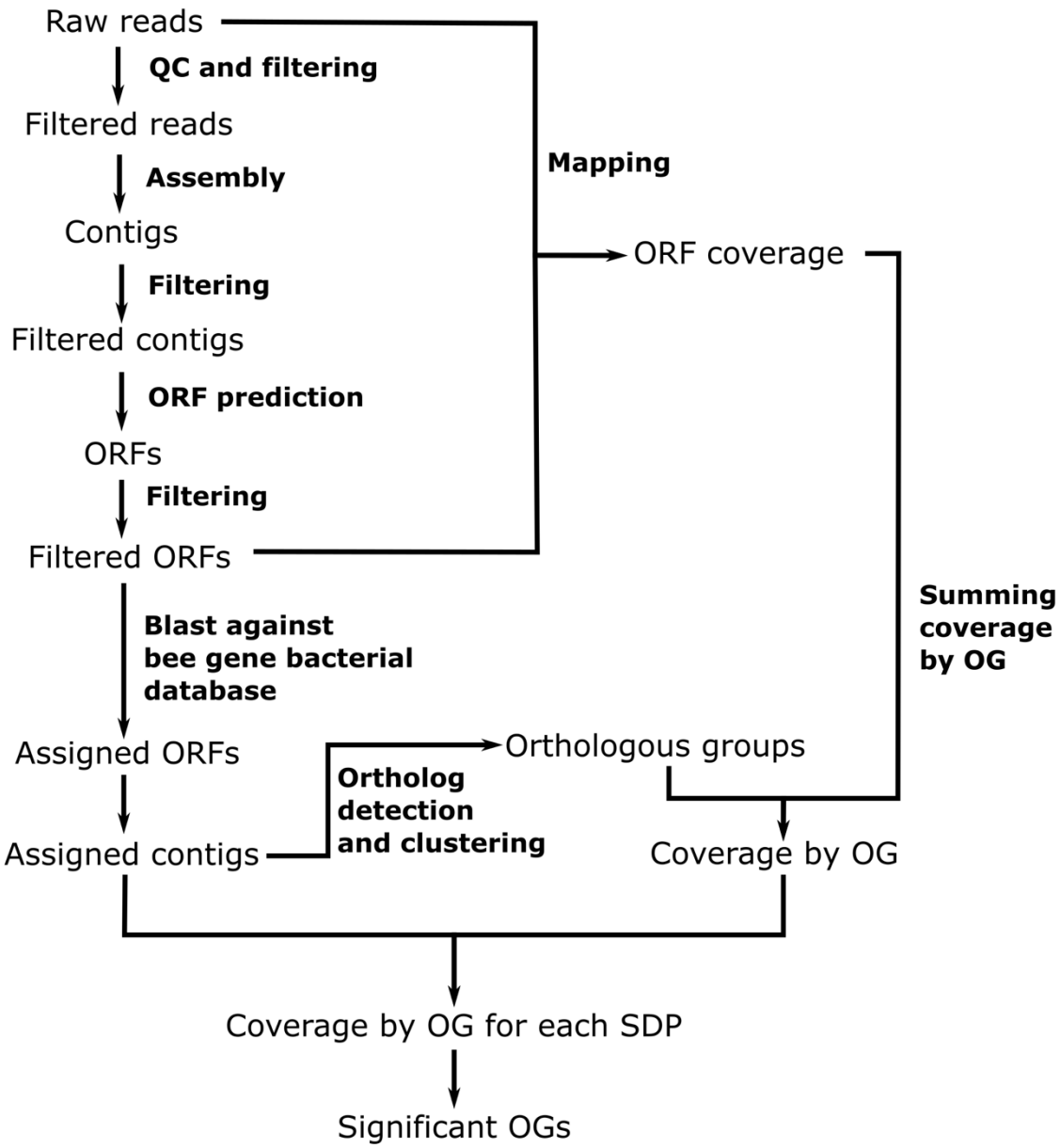
Supplementary Figure 5. Absolute abundance of the different species detected in the samples approximated by the normalized estimated number of 16S rRNA gene copies. Asterisks indicate q-values < 0.05.



Supplementary Figure 6. Pearson correlation coefficient and non-corrected p-value for the correlation of the percentage of polymorphic sites per species with the coverage of the species normalized by the bacterial load for each sample.



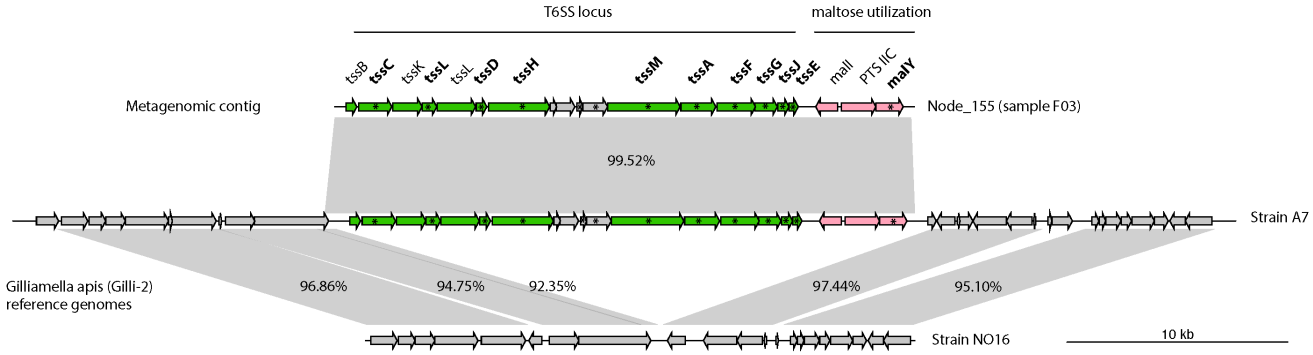
Supplementary Figure 7. Principal Coordinate Analyses of the Jaccard distance between all samples for the different species based on proportions of shared polymorphic sites. The color scheme shows information on the sampling sites of the different samples. The connecting lines indicates the two samples from the same colony.



726

727 **Supplementary Figure 8.** Workflow used for the functional gene content analysis.

728



729

730 **Supplementary Figure 9**. Genomic island of Gilli_2 encoding T6SS genes enriched in forager

731 versus nurse samples. The genomic island is only present in strain A7 of Gilli_2 in our genome

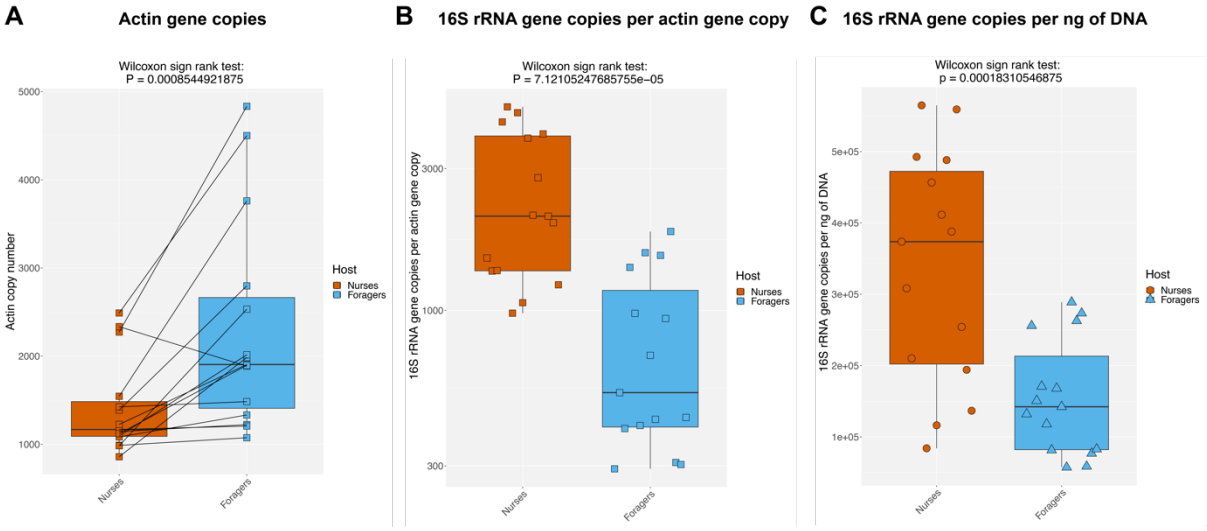
732 database and is integrated in a conserved genomic region. Genes in green indicate T6SS

733 genes, genes in pink depict strain-specific genes beloning to the same genomic island, and

734 genes in grey are other other genes present in the flanking regions of genomes of Gilli_2.

735 Genes enriched in forager samples are labeled in bold and flagged with an asterisk.

736



737

738 **Supplementary Figure 10.** (A) Actin gene copies in 10ng of each sample. (B) 16S rRNA gene

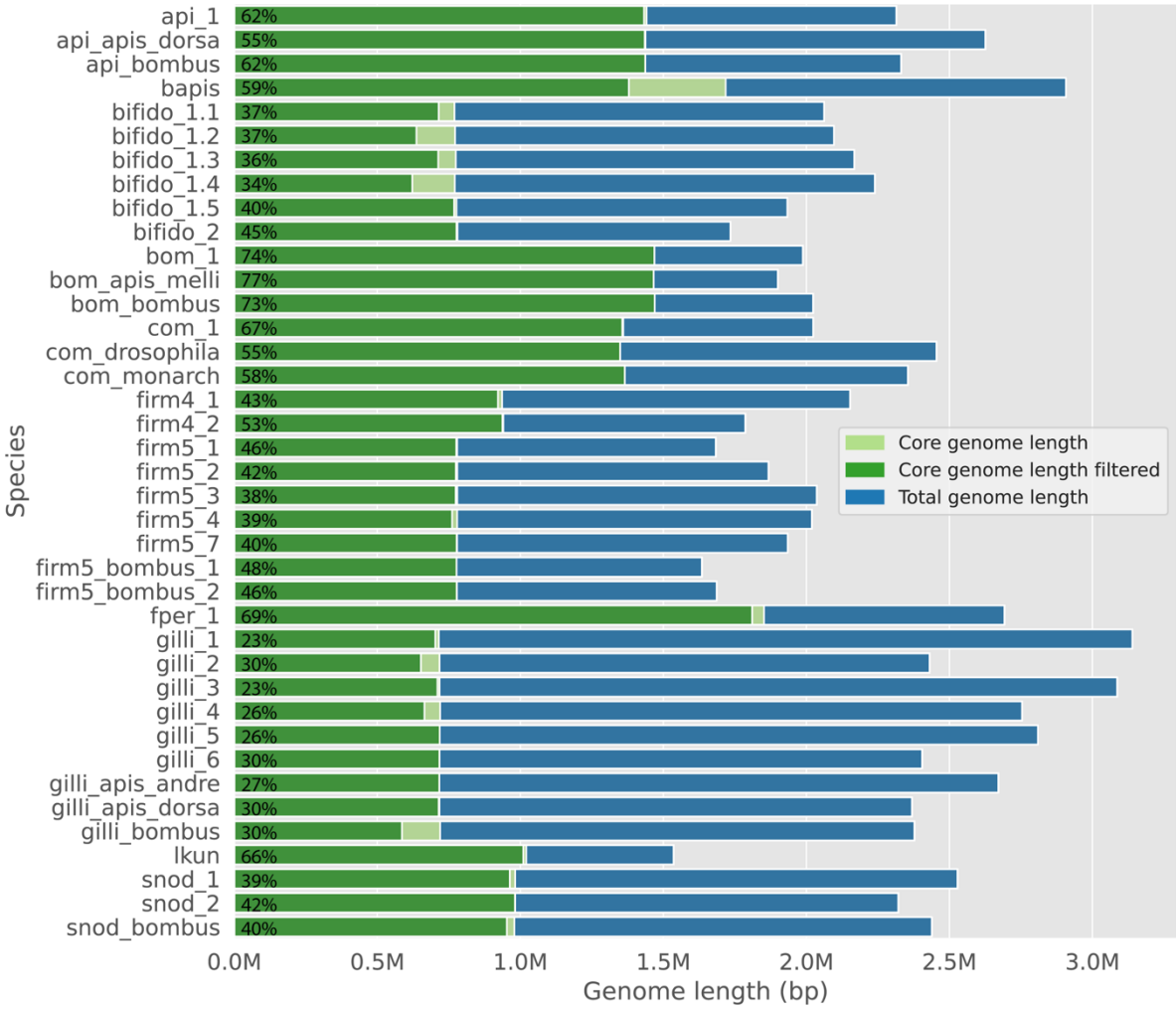
739 copy number per actin gene copy per sample. (C) rRNA gene copy number per actin gene

740 copy per ng of total DNA. The normalization with actin gene copies amplifies the differences

741 between the two groups, as there are fewer copies in foragers than in nurses.

742

743



744

745 **Supplementary Figure 11.** Cumulative length of all single-copy core genes (Core genome

746 length, in green) relative to the total genome length of the reference strain (Total genome

747 length, in blue) of each species. Reference strain is the strain used in the reduced reference

748 database. The filtered core genome length represents the portion of the core genes which

749 were used for the SNV analysis as they had >10x read coverage across all samples.

750

751

752

753 **Supplementary Table 1.** Sample collection information.

Sample name	Collection date	Treatment	Location	Number of guts
N01	25.06.2019	Larvae-feeding nurses	Lausanne	19
N02	25.06.2019	Larvae-feeding nurses	Lausanne	18
N03	25.06.2019	Larvae-feeding nurses	Lausanne	20
N04	26.07.2019	Larvae-feeding nurses	Liebefeld	20
N06	26.07.2019	Larvae-feeding nurses	Liebefeld	20
N07	26.07.2019	Larvae-feeding nurses	Liebefeld	20
N08	30.07.2019	Larvae-feeding nurses	Yens	20
N09	30.07.2019	Larvae-feeding nurses	Yens	20
N10	30.07.2019	Larvae-feeding nurses	Yens	20
N11	06.08.2019	Larvae-feeding nurses	Cugy	20
N12	06.08.2019	Larvae-feeding nurses	Cugy	20
N13	06.08.2019	Larvae-feeding nurses	Cugy	20
N14	27.08.2019	Larvae-feeding nurses	Vesancy	20
N15	27.08.2019	Larvae-feeding nurses	Vesancy	20
N16	27.08.2019	Larvae-feeding nurses	Vesancy	20
F01	25.06.2019	Pollen foragers	Lausanne	20
F02	25.06.2019	Pollen foragers	Lausanne	20
F03	25.06.2019	Pollen foragers	Lausanne	20
F04	26.07.2019	Pollen foragers	Liebefeld	20
F06	26.07.2019	Pollen foragers	Liebefeld	20
F07	26.07.2019	Pollen foragers	Liebefeld	20
F08	30.07.2019	Pollen foragers	Yens	20
F09	30.07.2019	Pollen foragers	Yens	20

F10	30.07.2019	Pollen foragers	Yens	20
F11	06.08.2019	Pollen foragers	Cugy	11
F12	06.08.2019	Pollen foragers	Cugy	16
F13	06.08.2019	Pollen foragers	Cugy	10
F14	27.08.2019	Pollen foragers	Vesancy	20
F15	27.08.2019	Pollen foragers	Vesancy	20
F16	27.08.2019	Pollen foragers	Vesancy	20

756 **Supplementary Table 2.** Table of all strain genomes included in the reference database, their
757 Species/SDP abbreviation used in this study, their corresponding full species name (if existing)
758 and the host where they were isolated. Genomes used as representative in the reduced
759 database are marked by 'y' in the last column.

Locus_tag	Strain_name	Phylotype	Species/SDP abbreviation	Species	Host	Reduced database
C4S75	wkB309	*Apibacter*	Api_1	*Apibacter sp.*	Apis cerana	n
Ga0307799	ESL0404	*Apibacter*	Api_1	*Apibacter sp.*	Apis cerana	n
Ga0312307	ESL0432	*Apibacter*	Api_1	*Apibacter sp.*	Apis cerana	y
C4S76	wkB180	*Apibacter*	Api_apis_dorsa	*Apibacter adventoris*	Apis dorsata	y
C4S77	wkB301	*Apibacter*	Api_apis_dorsa	*Apibacter adventoris*	Apis dorsata	n
Ga0061079	R-53146	*Apibacter*	Api_bombus	*Apibacter mensalis*	Bombus lapidarius	y
BBC0122	BBC0122	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	y
BBC0178	BBC0178	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
BBC0244	BBC0244	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S82	M0192	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S80	M0177	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S83	W8122	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S84	W8098	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S85	M0187	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S89	B10834G6	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S90	W8097	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3S98	B10834H15	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3U94	W8125	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n

H3V12	W8167	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3V13	M0280	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3V16	W8099	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
H3V17	M0283	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
PEB0122	PEB0122	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
PEB0149	PEB0149	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
PEB0150	PEB0150	*Bartonella*	Bapis	*Bartonella apis*	Apis mellifera	n
Ga0072395	Bin2	*Bifidobacterium*	Bifido_1.1	*bifidobacterium apousia*	Apis mellifera	n
H3T81	M0109	*Bifidobacterium*	Bifido_1.1	*bifidobacterium apousia*	Apis mellifera	n
H3T92	W8112	*Bifidobacterium*	Bifido_1.1	*bifidobacterium apousia*	Apis mellifera	y
H3U88	W8120	*Bifidobacterium*	Bifido_1.1	*bifidobacterium apousia*	Apis mellifera	n
Ga0072396	Bin7	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
H3S97	B14384H11	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
H3T86	W8113	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
H3U87	W8101	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
H3U89	W8104	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	y

H3U98	W8116	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
H3V04	M0353	*Bifidobacterium*	Bifido_1.2	*Bifidobacterium choladohabitans*	Apis mellifera	n
BAST	PRL2011	*Bifidobacterium*	Bifido_1.3	*Bifidobacterium asteroides*	Apis mellifera	y
Ga0133549	ESL0170	*Bifidobacterium*	Bifido_1.3	*Bifidobacterium asteroides*	Apis mellifera	n
H3S92	W8130	*Bifidobacterium*	Bifido_1.3	*Bifidobacterium asteroides*	Apis mellifera	n
H3S93	W8109	*Bifidobacterium*	Bifido_1.3	*Bifidobacterium asteroides*	Apis mellifera	n
H3T89	W8114	*Bifidobacterium*	Bifido_1.3	*Bifidobacterium asteroides*	Apis mellifera	n
Ga0072401	Hma3	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	n
Ga0133551	ESL0198	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	n
H3T90	W8117	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	n
H3T91	B14448H7	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	y
H3T93	M0404	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	n
H3V06	W6238G3	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium polysaccharolyticum*	Apis mellifera	n
Ga0133552	ESL0199	*Bifidobacterium*	Bifido_1.4	*Bifidobacterium sp.*	Apis mellifera	n

Ga0133553	ESL0200	*Bifidobacterium*	Bifido_1.5	*Bifidobacterium sp.*	Apis mellifera	y
Ga0133550	ESL0197	*Bifidobacterium*	Bifido_2	*Bifidobacterium indicum/coryneforme*	Apis mellifera	n
Ga0072398	Bma6	*Bifidobacterium*	Bifido_2	*Bifidobacterium indicum/coryneforme*	Apis mellifera	n
BINDI	LMG 11587	*Bifidobacterium*	Bifido_2	*Bifidobacterium indicum*	Insect	y
BCOR	LMG 18911	*Bifidobacterium*	Bifido_2	*Bifidobacterium coryneforme*	Apis mellifera	n
H3V08	M0307	*Bifidobacterium*	Bifido_2	*Bifidobacterium sp.*	Apis mellifera	n
Ga0098206	DSM 22766	*Bifidobacterium*	Bifido_bombus	*Bifidobacterium actinocoloniiforme*	Bumblebee digestive tract	n
Ga0056940	DSM 19703	*Bifidobacterium*	Bifido_bombus	*Bifidobacterium bombi*	Bumblebee digestive tract	n
Ga0057557	DSM 22767	*Bifidobacterium*	Bifido_bombus	*Bifidobacterium bohemicum*	Bumblebee digestive tract	n
Ga0312305	ESL0447	*Bifidobacterium*	Bifido_1_cerana	*Bifidobacterium sp.*	Apis cerana	n
Ga0372762	ESL0368	*Bombella*	Bom_1	*Bombella sp.*	Apis mellifera	y
Ga0216357	ESL0385	*Bombella*	Bom_apis_melli	*Bombella sp.*	Apis mellifera	y
Ga0308518	R-52487	*Bombella*	Bom_bombus	*Bombella intestini*	Bombus lapidarius	y
D9V35	AMU001	*Commensalibacter*	Com_1	*Commensalibacter*	Apis mellifera	y
H3T57	W6292M3	*Commensalibacter*	Com_1	*Commensalibacter*	Apis mellifera	n
CIN	A911	*Commensalibacter*	Com_drosophila	*Commensalibacter intestini*	Drosophila melagenogaster	y
Ga0248239	MX01	*Commensalibacter*	Com_monarch	*Commensalibacter*	Monarch butterfly	y
Ga0070888	Hon2	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis mellifera	n

Ga0227357	ESL0294	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis mellifera	n
Ga0227358	ESL0295	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis mellifera	n
Ga0227360	ESL0394	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis mellifera	n
Ga0326456	ESL0449	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis cerana	y
H3T38	W8088	*Bombilactobacillus*	Firm4_1	*Bombilactobacillus mellis*	Apis mellifera	n
Ga0072399	Bin4	*Bombilactobacillus*	Firm4_2	*Bombilactobacillus mellifer*	Apis mellifera	y
H3R26	W8092	*Bombilactobacillus*	Firm4_2	*Bombilactobacillus mellifer*	Apis mellifera	n
Ga0133563	ESL0185	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	y
Ga0061073	LMG 26964	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
Ga0072402	ESL0353	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
Ga0226842	ESL0263	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
H3R21	W8172	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
H3R25	M0390	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
H3T48	M0403	*Lactobacillus*	Firm5_1	*Lactobacillus apis*	Apis mellifera	n
Ga0133561	ESL0183	*Lactobacillus*	Firm5_2	*Lactobacillus helsingborgensis*	Apis mellifera	y
Ga0072400	ESL0354	*Lactobacillus*	Firm5_2	*Lactobacillus helsingborgensis*	Apis mellifera	n
LACWKB8	wkB8	*Lactobacillus*	Firm5_2	*Lactobacillus helsingborgensis*	Apis mellifera	n
H3T47	W8093	*Lactobacillus*	Firm5_2	*Lactobacillus helsingborgensis*	Apis mellifera	n
Ga0133562	ESL0184	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	y

Ga0072404	ESL0350	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	n
Ga0225908	ESL0259	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	n
Ga0226840	ESL0260	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	n
Ga0227359	ESL0393	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	n
H3R22	W8171	*Lactobacillus*	Firm5_3	*Lactobacillus melliventris*	Apis mellifera	n
Ga0072403	ESL0352	*Lactobacillus*	Firm5_4	*Lactobacillus kimbladii*	Apis mellifera	n
Ga0070887	ESL0351	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	n
LACWKB10	wkB10	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	n
Ga0226839	ESL0261	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	n
Ga0133564	ESL0186	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	y
H3T43	M0396	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	n
H3U49	W8174	*Lactobacillus*	Firm5_4	*Lactobacillus kullabergensis*	Apis mellifera	n
Ga0307804	ESL0409	*Lactobacillus*	Firm5_7	*Lactobacillus panisapium*	Apis cerana	n
Ga0311394	ESL0411	*Lactobacillus*	Firm5_7	*Lactobacillus panisapium*	Apis cerana	n
Ga0312303	ESL0413	*Lactobacillus*	Firm5_7	*Lactobacillus panisapium*	Apis cerana	y
Ga0312304	ESL0416	*Lactobacillus*	Firm5_7	*Lactobacillus panisapium*	Apis cerana	n
Ga0226852	ESL0228	*Lactobacillus*	Firm5_bombus_1	*Lactobacillus bombicola*	Bombus bohemicus	y
Ga0226843	ESL0233	*Lactobacillus*	Firm5_bombus_2	*Lactobacillus bombicola*	Bombus terrestris	n

Ga0226847	ESL0230	*Lactobacillus*	Firm5_bombus_2	*Lactobacillus bombicola*	Bombus pascuorum	y
Ga0061072	R-53102	*Lactobacillus*	Firm5_bombus_2	*Lactobacillus bombicola*	Bombus lapidarius	n
Ga0077910	ESL0034	*Frischella*	Fper_1	*Frischella perrara*	Apis mellifera	y
Ga0133554	ESL0167	*Frischella*	Fper_1	*Frischella perrara*	Apis mellifera	n
Ga0418777	Fper_2	*Frischella*	Fper_2	*Frischella sp.*	Apis cerana	n
Ga0133559	ESL0178	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
GAPWK	wkB1	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	y
B6D22	AM4	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
A9G14	P54G	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
A9G17	wkB7	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
B6D03	N-G5	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
B6D19	N-22	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
B6D08	A-7-12	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
B5S40	A-1-24	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
H3S73	B14384G12	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
H3T96	W8136	*Gilliamella*	Gilli_1	*Gilliamella apicola*	Apis mellifera	n
Ga0133555	ESL0169	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	y
Ga0133556	ESL0172	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6C99	N-G2	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6D18	A7	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
A9G16	P83G	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
A9G19	M1-2G	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n

A9G15	P62G	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6C96	NO15	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6D14	N-G4	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6D17	N-G1	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6D21	AM1	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
B6C88	NO16	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
H3U96	M0364	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
H3T80	W8145	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
H3T56	W8123	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
H3T83	M0320	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
H3U97	W8128	*Gilliamella*	Gilli_2	*Gilliamella apis*	Apis mellifera	n
Ga0133557	ESL0177	*Gilliamella*	Gilli_3	*Gilliamella apis*	Apis mellifera	y
Ga0133558	ESL0182	*Gilliamella*	Gilli_3	*Gilliamella apis*	Apis mellifera	n
Ga0307802	ESL0441	*Gilliamella*	Gilli_4	*Gilliamella sp.*	Apis cerana	y
A9G10	wkB308	*Gilliamella*	Gilli_4	*Gilliamella sp.*	Apis cerana	n
Ga0307800	ESL0405	*Gilliamella*	Gilli_5	*Gilliamella sp.*	Apis cerana	y
Ga0307803	ESL0443	*Gilliamella*	Gilli_6	*Gilliamella sp.*	Apis cerana	y
A9G07	wkB72	*Gilliamella*	Gilli_6	*Gilliamella sp.*	Apis cerana	n
A9G08	wkB195	*Gilliamella*	Gilli_6	*Gilliamella sp.*	Apis cerana	n
A9G09	wkB292	*Gilliamella*	Gilli_apis_andre	*Gilliamella sp.*	Apis cerana	n
A9G03	wkB171	*Gilliamella*	Gilli_apis_andre	*Gilliamella sp.*	Apis andreniformis	y
A9G11	wkB108	*Gilliamella*	Gilli_apis_dorsa	*Gilliamella sp.*	Apis dorsata	n
A9G12	wkB112	*Gilliamella*	Gilli_apis_dorsa	*Gilliamella sp.*	Apis dorsata	n
A9G13	wkB178	*Gilliamella*	Gilli_apis_dorsa	*Gilliamella sp.*	Apis dorsata	y

A9G23	App4-10	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus appositus	n
A9G24	App6-5	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus appositus	n
A9G29	Fer2-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus fervidus	n
Ga0055041	wkB11	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus bimaculatus	n
A9G30	Fer4-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus fervidus	n
A9G31	Gris1-4	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus griseocollis	n
A9G32	Gris3-2	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus griseocollis	n
A9G35	Choc5-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus impatiens	n
A9G37	GilliExp13	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus impatiens	n
A9G38	Imp1-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus impatiens	n
A9G39	Imp1-6	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus impatiens	n
A9G41	Nev5-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus nevadensis	n
A9G42	Nev6-6	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus nevadensis	n
A9G43	Occ3-1	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus occidentalis	n
A9G44	Occ4-3	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus occidentalis	n
A9G47	WF3-4	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus pensylvanicus	n
Ga0227298	ESL0232	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus terrestris	n
Ga0227301	ESL0250	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus pascuorum	n
Ga0227302	ESL0254	*Gilliamella*	Gilli_bombus	*Gilliamella sp.*	Bombus pascuorum	n
Ga0061080	R-53144	*Gilliamella*	Gilli_bombus	*Gilliamella intestini*	Soil	n
Ga0098617	Fhon2	*Apilactobacillus*	Lkun	*Apilactobacillus kunkeei*	Apis mellifera	n

Ga0308523	Fhon13	*Apilactobacillus*	Lkun	*Apilactobacillus kunkeei*	Apis mellifera	n
H3T37	M0345	*Apilactobacillus*	Lkun	*Apilactobacillus kunkeei*	Apis mellifera	n
Ga0326500	ESL0196	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
SALWKB2	wkB2	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
CPT77	E1	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGH94	Aw-20	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGH95	Aw-18	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGH96	A12	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGH97	A-10-12	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI00	A-1-12	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI01	A2	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI08	N-W7	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI09	N9	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI12	N-S3	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BGI13	N-S4	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BHC42	wkB332	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BHC47	PEB0171	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BHC50	wkB339	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BHC52	MS1-3	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
Ga0227305	ESL0323	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
Ga0227306	ESL0324	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n

Ga0227304	ESL0304	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3U73	W8124	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3T75	W8135	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3T79	M0118	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3U82	W8132	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3V10	W6238H14	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3V01	M0351	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
H3V11	W8158	*Snodgrassella*	Snod_1	*Snodgrassella alvi*	Apis mellifera	n
BHC53	wkB237	*Snodgrassella*	Snod_2	*Snodgrassella sp.*	Apis andreniformis	n
BHC43	wkB298	*Snodgrassella*	Snod_2	*Snodgrassella sp.*	Apis cerana	n
BHC54	wkB273	*Snodgrassella*	Snod_2	*Snodgrassella sp.*	Apis florea	n
BHC44	Nev4-2	*Snodgrassella*	Snod_bombus	*Snodgrassella sp.*	Bombus nevadensis	n
BHC46	Ruf1-X	*Snodgrassella*	Snod_bombus	*Snodgrassella alvi*	Bombus rufocinctus	n
BHC48	Occ4-2	*Snodgrassella*	Snod_bombus	*Snodgrassella communis*	Bombus occidentalis	n
BHC57	HK3	*Snodgrassella*	Snod_bombus	*Snodgrassella sp.*	Bombus pensylvanicus	n
BGI32	App2-2	*Snodgrassella*	Snod_bombus	*Snodgrassella sp.*	Bombus appositus	n
BGI34	App6-4	*Snodgrassella*	Snod_bombus	*Snodgrassellla sp.*	Bombus appositus	n
BGI36	Fer2-2	*Snodgrassella*	Snod_bombus	*Snodgrassella communis*	Bombus fervidus	n
SALWKB12	wkB12	*Snodgrassella*	Snod_bombus	*Snodgrassella communis*	Bombus bimaculatus	n
Ga0227310	ESL0253	*Snodgrassella*	Snod_bombus	*Snodgrassella sp.*	Bombus pascuorum	n

762 **Supplementary Table 3.** Proportions of reads mapping to species isolated in non-*A. mellifera* hosts.

763 The proportions are computed as the number of reads mapping to the species, divided by the

764 number of reads mapping to the same phylotype for this sample.

Species	**Sample**	**Proportion of reads (%) mapping out of the reads mapping to the same phylotype for this sample**
Firm5_7	N02	0.47
	N04	0.68
	N06	0.37
	N10	0.12
	N13	3.1
Gilli_bombus	N10	0.16
	N14	0.15
Gilli_apis_dorsa	F11	0.84
	F16	0.92

765

766 **Supplementary Table 4**. Matrix of the counts of the OGs annotated as H category significantly

767 enriched in the forager samples compared to the al OGs and all COG categories.

	"H" COG category	**All COG category**
All OGs	71	1779
OGs significantly enriched in foragers	18	63

768

769

770 ## References

771 1. Keeney KM, Yurist-Doutsch S, Arrieta MC, Finlay BB: **Effects of antibiotics on human**
772 **microbiota and subsequent disease.** *Annu Rev Microbiol* 2014, **68:**217-235.
773 2. Martin-Platero AM, Cleary B, Kauffman K, Preheim SP, McGillicuddy DJ, Alm EJ, Polz MF:
774 **High resolution time series reveals cohesive but short-lived communities in coastal**
775 **plankton.** *Nat Commun* 2018, **9:**266.
776 3. Teeling H, Fuchs BM, Bennke CM, Kruger K, Chafee M, Kappelmann L, Reintjes G, Waldmann
777 J, Quast C, Glockner FO, et al: **Recurring patterns in bacterioplankton dynamics during**
778 **coastal spring algae blooms.** *Elife* 2016, **5:**e11888.
779 4. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang
780 S, Mann AJ, Waldmann J, et al: **Substrate-controlled succession of marine bacterioplankton**
781 **populations induced by a phytoplankton bloom.** *Science* 2012, **336:**608-611.
782 5. Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, Knight R, Manjurano A,
783 Changalucha J, Elias JE, et al: **Seasonal cycling in the gut microbiome of the Hadza hunter-**
784 **gatherers of Tanzania.** *Science* 2017, **357:**802-806.
785 6. Sommer F, Stahlman M, Ilkayeva O, Arnemo JM, Kindberg J, Josefsson J, Newgard CB,
786 Frobert O, Backhed F: **The Gut Microbiota Modulates Energy Metabolism in the**
787 **Hibernating Brown Bear Ursus arctos.** *Cell Rep* 2016, **14:**1655-1661.
788 7. Maurice CF, Knowles SC, Ladau J, Pollard KS, Fenton A, Pedersen AB, Turnbaugh PJ: **Marked**
789 **seasonal variation in the wild mouse gut microbiota.** *ISME J* 2015, **9:**2423-2434.
790 8. Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y: **Seasonal variation**
791 **in human gut microbiome composition.** *PLoS One* 2014, **9:**e90731.
792 9. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S: **Multi-kingdom**
793 **ecological drivers of microbiota assembly in preterm infants.** *Nature* 2021, **591:**633-638.

794 10. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong
795 H, et al: **Dynamics and Stabilization of the Human Gut Microbiome during the First Year of**
796 **Life.** *Cell Host Microbe* 2015, **17:**690-703.
797 11. Patnode ML, Guruge JL, Castillo JJ, Couture GA, Lombard V, Terrapon N, Henrissat B, Lebrilla
798 CB, Gordon JI: **Strain-level functional variation in the human gut microbiota based on**
799 **bacterial binding to artificial food particles.** *Cell Host Microbe* 2021, **29:**664-673 e665.
800 12. Barboza M, Sela DA, Pirim C, Locascio RG, Freeman SL, German JB, Mills DA, Lebrilla CB:
801 **Glycoprofiling bifidobacterial consumption of galacto-oligosaccharides by mass**
802 **spectrometry reveals strain-specific, preferential consumption of glycans.** *Appl Environ*
803 *Microbiol* 2009, **75:**7319-7325.
804 13. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ: **Predicting and**
805 **manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta.**
806 *Science* 2013, **341:**295-298.
807 14. Bron PA, van Baarlen P, Kleerebezem M: **Emerging molecular insights into the interaction**
808 **between probiotics and the host intestinal mucosa.** *Nat Rev Microbiol* 2011, **10:**66-78.
809 15. Goyal A, Bittleston LS, Leventhal GE, Lu L, Cordero OX: **Interactions between strains govern**
810 **the eco-evolutionary dynamics of microbial communities.** *Elife* 2022, **11**.
811 16. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, Cordero OX: **Strain-level**
812 **diversity drives alternative community types in millimetre-scale granular biofilms.** *Nat*
813 *Microbiol* 2018, **3:**1295-1303.
814 17. Lou YC, Olm MR, Diamond S, Crits-Christoph A, Firek BA, Baker R, Morowitz MJ, Banfield JF:
815 **Infant gut strain persistence is associated with maternal origin, phylogeny, and traits**
816 **including surface adhesion and iron acquisition.** *Cell Rep Med* 2021, **2:**100393.
817 18. Wolff R, Shoemaker W, Garud N: **Ecological Stability Emerges at the Level of Strains in the**
818 **Human Gut Microbiome.** *mBio* 2023, **14:**e0250222.

819 19. Bonilla-Rosso G, Engel P: **Functional roles and metabolic niches in the honey bee gut**
820 **microbiota.** *Curr Opin Microbiol* 2018, **43:**69-76.
821 20. Kwong WK, Moran NA: **Gut microbial communities of social bees.** *Nat Rev Microbiol* 2016,
822 **14:**374-384.
823 21. Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, Jaron KS, van der
824 Meer JR, Robinson-Rechavi M, Sentchilo V, et al: **Genomic changes underlying host**
825 **specialization in the bee gut symbiont Lactobacillus Firm5.** *Mol Ecol* 2019, **28:**2224-2237.
826 22. Ellegaard KM, Engel P: **Genomic diversity landscape of the honey bee gut microbiota.** *Nat*
827 *Commun* 2019, **10:**446.
828 23. Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vasquez A: **Extensive intra-**
829 **phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut.** *BMC*
830 *Genomics* 2015, **16:**284.
831 24. Segers FH, Kesnerova L, Kosoy M, Engel P: **Genomic changes associated with the**
832 **evolutionary transition of an insect gut symbiont into a blood-borne pathogen.** *ISME J*
833 2017, **11:**1232-1244.
834 25. Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA: **Metabolism of Toxic**
835 **Sugars by Strains of the Bee Gut Symbiont Gilliamella apicola.** *MBio* 2016, **7**.
836 26. Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA: **Division of**
837 **labor in honey bee gut microbiota for plant polysaccharide digestion.** *Proc Natl Acad Sci U*
838 *S A* 2019, **116:**25909-25916.
839 27. Brodschneider R, Crailsheim, K: **Nutrition and health in honey bees.** *Apidologie* 2010,
840 **41:**278-294.
841 28. Seeley TD: *The Wisdom of the Hive: the social physiology of honey bee colonies.* Harvard
842 University Press; 1996.
843 29. Smith CR, Toth AL, Suarez AV, Robinson GE: **Genetic and genomic analyses of the division of**
844 **labour in insect societies.** *Nat Rev Genet* 2008, **9:**735-748.

845 30. Robinson GE: **Genomics and integrative analyses of division of labor in honeybee colonies.**
846 *Am Nat* 2002, **160 Suppl 6:**S160-172.
847 31. Whitfield CW, Ben-Shahar Y, Brillet C, Leoncini I, Crauser D, Leconte Y, Rodriguez-Zas S,
848 Robinson GE: **Genomic dissection of behavioral maturation in the honey bee.** *Proc Natl*
849 *Acad Sci U S A* 2006, **103:**16068-16075.
850 32. Jones JC, Fruciano C, Marchant J, Hildebrand F, Forslund S, Bork P, Engel P, Hughes WOH:
851 **The gut microbiome is associated with behavioural task in honey bees.** *Insectes Soc* 2018,
852 **65:**419-429.
853 33. Kesnerova L, Emery O, Troilo M, Liberti J, Erkosar B, Engel P: **Gut microbiota structure**
854 **differs between honeybees in winter and summer.** *ISME J* 2020, **14:**801-814.
855 34. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, Hingamp P, Alves R,
856 Costea PI, Coelho LP, et al: **Microbial abundance, activity and population genomic profiling**
857 **with mOTUs2.** *Nat Commun* 2019, **10:**1014.
858 35. Ellegaard KM, Suenami S, Miyazaki R, Engel P: **Vast Differences in Strain-Level Diversity in**
859 **the Gut Microbiota of Two Closely Related Honey Bee Species.** *Curr Biol* 2020, **30:**2520-
860 2531 e2527.
861 36. Kwong WK, Medina LA, Koch H, Sing KW, Soh EJY, Ascher JS, Jaffe R, Moran NA: **Dynamic**
862 **microbiome evolution in social bees.** *Sci Adv* 2017, **3:**e1600513.
863 37. Ansorge R, Romano S, Sayavedra L, Porras MAG, Kupczok A, Tegetmeyer HE, Dubilier N,
864 Petersen J: **Functional diversity enables multiple symbiont strains to coexist in deep-sea**
865 **mussels.** *Nat Microbiol* 2019, **4:**2487-2497.
866 38. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin
867 J, Pati A, Bushnell B, et al: **Genome-wide selective sweeps and gene-specific sweeps in**
868 **natural bacterial populations.** *ISME J* 2016, **10:**1589-1601.

869 39. Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, Armanini F, Manghi P,
870 Bonham K, Zolfo M, et al: **The Prevotella copri Complex Comprises Four Distinct Clades**
871 **Underrepresented in Westernized Populations.** *Cell Host Microbe* 2019, **26:**666-679 e667.
872 40. Zhu A, Sunagawa S, Mende DR, Bork P: **Inter-individual differences in the gene content of**
873 **human gut bacterial species.** *Genome Biol* 2015, **16:**82.
874 41. Van Rossum T, Ferretti P, Maistrenko OM, Bork P: **Diversity within species: interpreting**
875 **strains in microbiomes.** *Nat Rev Microbiol* 2020, **18:**491-506.
876 42. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J,
877 Wijmenga C, Zhernakova A, Segal E: **Structural variation in the gut microbiome associates**
878 **with host health.** *Nature* 2019, **568:**43-48.
879 43. De Filippis F, Pasolli E, Tett A, Tarallo S, Naccarati A, De Angelis M, Neviani E, Cocolin L,
880 Gobbetti M, Segata N, Ercolini D: **Distinct Genetic and Functional Traits of Human Intestinal**
881 **Prevotella copri Strains Are Associated with Different Habitual Diets.** *Cell Host Microbe*
882 2019, **25:**444-453 e443.
883 44. Carrow HC, Batachari LE, Chu H: **Strain diversity in the microbiome: Lessons from**
884 **Bacteroides fragilis.** *PLoS Pathog* 2020, **16:**e1009056.
885 45. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, Sinha T, Lifelines Cohort S,
886 Segal E, Weersma RK, et al: **The long-term genetic stability and individual specificity of the**
887 **human gut microbiome.** *Cell* 2021, **184:**2302-2315 e2312.
888 46. Greenblum S, Carr R, Borenstein E: **Extensive strain-level copy-number variation across**
889 **human gut microbiome species.** *Cell* 2015, **160:**583-594.
890 47. Wu J, Lang H, Mu X, Zhang Z, Su Q, Hu X, Zheng H: **Honey bee genetics shape the strain-**
891 **level structure of gut microbiota in social transmission.** *Microbiome* 2021, **9:**225.
892 48. Li C, Tang M, Li X, Zhou X: **Community Dynamics in Structure and Function of Honey Bee**
893 **Gut Bacteria in Response to Winter Dietary Shift.** *mBio* 2022, **13:**e0113122.

894 49. Wu Y, Zheng Y, Wang S, Chen Y, Tao J, Chen Y, Chen G, Zhao H, Wang K, Dong K, et al:
895 **Genetic divergence and functional convergence of gut bacteria between the Eastern honey**
896 **bee Apis cerana and the Western honey bee Apis mellifera.** *J Adv Res* 2022, **37:**19-31.
897 50. Engel P, Martinson VG, Moran NA: **Functional diversity within the simple gut microbiota of**
898 **the honey bee.** *Proc Natl Acad Sci U S A* 2012, **109:**11002-11007.
899 51. Li Y, Leonard SP, Powell JE, Moran NA: **Species divergence in gut-restricted bacteria of**
900 **social bees.** *Proc Natl Acad Sci U S A* 2022, **119:**e2115013119.
901 52. Brochet S, Quinn A, Mars RAT, Neuschwander N, Sauer U, Engel P: **Niche partitioning**
902 **facilitates coexistence of closely related honey bee gut bacteria.** *Elife* 2021, **10**.
903 53. Kesnerova L, Mars RAT, Ellegaard KM, Troilo M, Sauer U, Engel P: **Disentangling metabolic**
904 **functions of bacteria in the honey bee gut.** *PLoS Biol* 2017, **15:**e2003467.
905 54. Kaval KG, Garsin DA: **Ethanolamine Utilization in Bacteria.** *mBio* 2018, **9**.
906 55. Elsayeh WAC, C.; Wright, G. A.: **B-Vitamins Influence the Consumption of Macronutrients in**
907 **Honey Bees.** *Front Sustain Food Syst* 2022, **6**.
908 56. Tsoy O, Ravcheev D, Mushegian A: **Comparative genomics of ethanolamine utilization.** *J*
909 *Bacteriol* 2009, **191:**7157-7164.
910 57. Martinson VG, Moy J, Moran NA: **Establishment of characteristic gut bacteria during**
911 **development of the honeybee worker.** *Appl Environ Microbiol* 2012, **78:**2830-2840.
912 58. Steele MI, Kwong WK, Whiteley M, Moran NA: **Diversification of Type VI Secretion System**
913 **Toxins Reveals Ancient Antagonism among Bee Gut Microbes.** *MBio* 2017, **8**.
914 59. Steele MI, Moran NA: **Evolution of Interbacterial Antagonism in Bee Gut Microbiota**
915 **Reflects Host and Symbiont Diversification.** *mSystems* 2021, **6**.
916 60. Powell JE, Leonard SP, Kwong WK, Engel P, Moran NA: **Genome-wide screen identifies host**
917 **colonization determinants in a bacterial gut symbiont.** *Proc Natl Acad Sci U S A* 2016,
918 **113:**13887-13892.

919 61. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence**
920 **data.** *Bioinformatics* 2014, **30:**2114-2120.
921 62. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S: **High throughput ANI analysis**
922 **of 90K prokaryotic genomes reveals clear species boundaries.** *Nat Commun* 2018, **9:**5114.
923 63. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons**
924 **dramatically improves orthogroup inference accuracy.** *Genome Biol* 2015, **16:**157.
925 64. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.**
926 *Bioinformatics* 2009, **25:**1754-1760.
927 65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
928 Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.**
929 *Bioinformatics* 2009, **25:**2078-2079.
930 66. Bremer H, Churchward G: **An examination of the Cooper-Helmstetter theory of DNA**
931 **replication in bacteria and its underlying assumptions.** *J Theor Biol* 1977, **69:**645-654.
932 67. Garrison EM, G.: **Haplotype-based variant detection from short-read sequencing.** *arXiv*
933 2012.
934 68. Garrison EK, Z. N.; Dawson, E. T.; Pedersen, B. S.; Prins, P.: **Vcflib and tools for processing**
935 **the VCF variant call format.** *bioRxiv* 2021.
936 69. **Picard Tools. Broad Institute.** [http://broadinstitute.github.io/picard/]
937 70. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA: **metaSPAdes: a new versatile**
938 **metagenomic assembler.** *Genome Res* 2017, **27:**824-834.
939 71. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene**
940 **recognition and translation initiation site identification.** *BMC Bioinformatics* 2010, **11:**119.
941 72. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P: **Fast**
942 **Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.**
943 *Mol Biol Evol* 2017, **34:**2115-2122.

944 73. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR,
945 Letunic I, Rattei T, Jensen LJ, et al: **eggNOG 5.0: a hierarchical, functionally and**
946 **phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.**
947 *Nucleic Acids Res* 2019, **47:**D309-D314.
948 74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol*
949 *Biol* 1990, **215:**403-410.
950 75. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons**
951 **dramatically improves orthogroup inference accuracy.** *Genome Biol* 2015, **16:**157.
952 76. Garsin DA: **Ethanolamine utilization in bacterial pathogens: roles and regulation.** *Nat Rev*
953 *Microbiol* 2010, **8:**290-295.
954